

## A door to be kept open

**Broad sanctions that could isolate Indian and Pakistani scientists from the West are a counter-productive response to the two nations' unwelcome arrival in the nuclear club. Sanctions should be used with care.**

Scientists in India and Pakistan have reacted indignantly to the discovery that some of their collaborations with international partners may be coming to an abrupt end. This is a result of sanctions imposed by the United States and others in the aftermath of the nuclear weapons tests conducted by both nations in May. They are entitled to be upset. But they can hardly be surprised. After all, international non-proliferation efforts that have sought with some success to restrict the spread of nuclear weapons are based on the concept that defiance should be punished. It is hard to see how such efforts can maintain any credibility if no punishment is now forthcoming.

Despite India's protestations (see page 513), the Western response so far has been relatively limited. The US government is only withdrawing funds from collaborations with agencies directly involved in the nuclear weapons and ballistic missiles programmes of both countries (India is the chief target, as such work had already ended under earlier economic sanctions against Pakistan). In some cases, the target of such sanctions is legitimate. But these agencies are also involved in many branches of research that have nothing to do with nuclear weapons or missiles.

The human cost of such actions can be considerable, even if they do not mark the end of scientific collaboration between the United States and the sub-continent. The State Department is showing some discretion; the law that automatically triggered sanctions when the Indian tests took place could be interpreted more harshly, requiring an end to all US-supported scientific collaboration. Amid allegations of a clamp-down on all scientists from the sub-continent seeking visas to travel to the United States, the State Department says that it is working out a new policy on the circumstances in which these can be issued.

As they do so, however, department officials should ask them-

selves who benefits from the curtailment of scientific contact or collaboration between India and Pakistan and the West. Such curtailment will do little to slow down the clandestine weapons programme of either country. But it will worsen the isolation of both at a time when greater engagement with the outside world is the best available hedge against a potentially cataclysmic military confrontation between them. Indeed the clearest beneficiary, perhaps, is the nationalist government of India, which started the whole crisis in the first place and would reap crude political gains from any moves to isolate Indian science and marginalize its scientists.

This contradiction affects not only scientific sanctions but all the others which the United States has put in place. The imposition of sanctions by the world's largest economy on the world's largest democracy was always likely to expose the limitations of such a strategy as an instrument of foreign policy. Indeed the US Congress has already admitted as much, exempting massive grain sales from the sanctions in order to please American farmers. The realization that the most economically significant sanctions cannot survive poses a new danger: that effective sanctions will be allowed to lapse, leaving in place a series of token measures, of which the restriction of scientific collaboration could well be one. That would be a ridiculous outcome.

Scientific collaboration between the Indian subcontinent and the West is an essential line of communication that both sides must now fight to preserve. At the very least, the National Academy of Sciences and the US scientific societies should press the State Department to maintain normal scientific relations as far as is possible within the letter of the sanctions law and, in particular, to maintain the right of scientists from the two countries to travel freely. Even non-proliferation efforts would benefit. □

## The right man in the right place

**Japan's choice of Akito Arima as its new minister of education is to be welcomed.**

In recent days, the eyes of the world have focused on the finance minister chosen by Japan's new prime minister, Keizo Obuchi, to try to pull Japan — and Asia — out of recession. But Obuchi's bold selection of reformist Akito Arima to head the conservative Ministry of Education, Science, Sports and Culture (Monbusho) is more likely to have a dramatic impact on Japan's long-term future — if his government stays in power long enough.

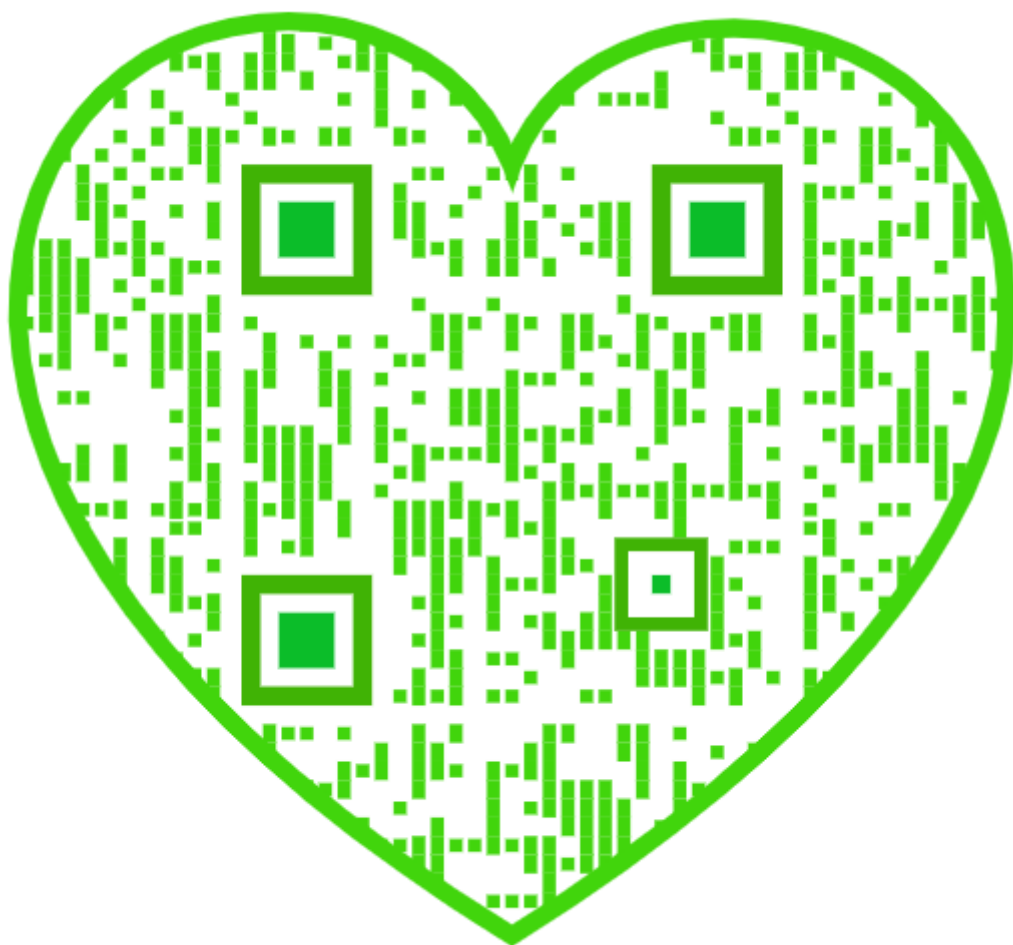
It is ironic that Arima, who recently served as president the Institute of Physical and Chemical Research (RIKEN), should be chosen. In 1991, when president of Tokyo University, Arima took the then minister of education, Yutaka Inoue, on a tour of the university's dilapidated science buildings, and won an apology from the minister for the sorry state of affairs. Now Arima has his own opportunity to bring about some real change.

Arima has championed the introduction of external reviews to assess research at universities and research institutes in Japan. He won new funds for renovating university infrastructure, and was a key player behind the 1996 Basic Science and Technology Law, which

has created a healthier environment for the funding of science in general. With his additional experience on government advisory committees, no one could be better qualified.

Hopefully, Arima will use his new post to push through a more widespread use of research assessment in Japan's university system, which has been resisting the introduction of such objective measures. But care must be taken to introduce sophisticated and fair means of assessment, rather than the simplistic measures of paper output that Arima has used in the past to sell the need for more evaluation.

Arima could also benefit Japan's universities profoundly if he could solve their chronic shortage of technicians. And his experience at RIKEN, under the Science and Technology Agency, makes him well qualified to take charge of the planned merger of the agency and the ministry of education. But it will require clever manoeuvring to ensure that Japan's science ends up better off as a whole, and that support for the pursuit of knowledge for its own sake in existing centres of excellence does not suffer at a time when so much emphasis is being placed on science's contribution to socioeconomic gain. □



# US Congress looks set to scuttle international fusion project

[WASHINGTON] The US Congress is poised to pass legislation seeking an end to US participation in the International Thermonuclear Experimental Reactor (ITER). The move is likely to lead to the collapse of a 13-year global effort to design and build a prototype fusion reactor.

The agreement to design ITER, intended to demonstrate sustained fusion in a magnetically confined plasma, expired on 21 July. Japan is said to feel that a three-year extension is contingent on the participation of all four parties — the United States, Russia, the European Union and Japan.

Russia's political interest is liable to wane if the United States withdraws. Officials say that the collaboration will therefore lie in ruins if, as seems likely, Congress passes an appropriations bill next month which pointedly advises the US Department of Energy to withdraw from the project.

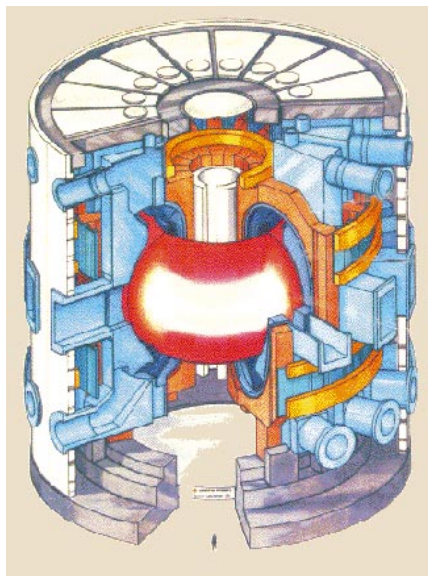
Even if this were to happen, however, there remains a possibility that Europe might decide to build a new fusion reactor either on its own, as it was planning before the ITER project was agreed on, or in a separate collaboration with Japan.

Energy department officials, backed by the US magnetic fusion research community, have been imploring members of the House energy and water appropriations subcommittee to withdraw the advice they inserted in their proposed 1999 appropriations bill.

The officials want the DoE to be allowed to sign the agreement, even if it cannot spend any money on ITER. They say this would at least allow the other partners, each of which has signed the extension agreement, to proceed. "We need to be able to sign, regardless of the money," says Anne Davies, head of the magnetic fusion programme at the DoE.

Joseph McDade (Republican, Pennsylvania), chair of the energy and water appropriations subcommittee, believes that several questions about ITER need a better answer before the agreement is extended. According to a letter which he sent to Federico Peña, then the energy secretary, on 11 June, these include questions about "whether this construction project will ever be started".

Last week, the energy and water appropriations subcommittees of both the House and the Senate were close to holding a conference at which, officials say, the language advising the DoE not to sign would have been confirmed. But they didn't get round to it before the Senate recessed on 31 July. This gives the ITER agreement a breathing space until their return in September, when a final



**Looking into ITER:** this may be all we ever see of it, if the United States backs out.

version of the appropriations bill must be agreed and passed into law.

James Sensenbrenner (Republican, Wisconsin), chairman of the House Science Committee, is to fly to Tokyo on 9 August to talk to senior Japanese officials about fusion research and other fields of US-Japanese research collaboration. Some officials and fusion scientists see this visit as a possible rescue mission for US participation in ITER.

In the cases of the International Space Station and the Large Hadron Collider, they note, Sensenbrenner has enjoyed making

international visits in which he has noisily banged the drum for American interests, scolded the Clinton administration for its alleged failure to protect these interests and then strongly supported both projects.

But sources close to the Wisconsin congressman are playing down such expectations in the case of ITER. During the five day trip, they say, he will try to soothe what are expected to be bruised Japanese feelings over the possible demise of the ITER agreement, find out if Japan is serious about a scaled down, incremental version of ITER, and look at possibilities for fusion research collaboration outside of ITER.

Although it is McDade — not Sensenbrenner — who is ready to kill the agreement, officials are nonetheless pinning their hopes on the latter. Asked if his trip could help save the agreement, Davies says: "I certainly hope so."

The last ITER Council meeting in Vienna amply demonstrated the fragile status of agreement. After meeting on 21 July — the date on which the six-year agreement to conduct an engineering design assessment of ITER expired — Japanese officials contended that the legal basis for the council had expired, as its terms require all four parties to agree on the proposed three-year extension.

The officials at Vienna then met informally on 22 July and agreed to keep talking informally until October, by which time the US situation should be resolved.

ITER research in the United States is continuing, at least until the end of the financial year on 30 September. "In terms of the tech-

## Retailer appointed UK minister for science

[LONDON] Lord Sainsbury, former chairman and chief executive of the large chain of supermarkets that bears his family's name, has been named as Britain's new science minister following a cabinet reshuffle last week.

Tony Blair, the prime minister, also announced that Sir Robert May, the government's chief scientist, will be given an office in the Cabinet Office, a move frequently demanded by critics since the Office of Science and Technology was moved to the Department of Trade and Industry (DTI) in 1995 (see *Nature* 376, 103; 1995).

Sainsbury was named last week as a new minister in the DTI, where he will serve under Peter Mandelson, a leading architect of the Labour Party's victory in last year's general election. He takes over the science

portfolio from John Battle, who retains his other responsibilities, including energy issues. Mandelson becomes the cabinet minister responsible for science and technology.

Sainsbury has long been an active promoter of scientific and technological causes, ranging from school education to support for research in plant pathology, through the activities of the Gatsby Foundation, the charitable trust on which he has settled much of his personal wealth.

The foundation's capital assets are said to be worth close to £500 million (US\$820 million). Projects it has backed include the creation of a plant research laboratory at the John Innes Centre in Norwich, and a unit for cognitive neuroscience at University College London.

David Dickson

nical work we are continuing with business as usual," says Charles Baker, head of the US ITER team based in San Diego, California.

But a senior scientist at one US fusion facility describes the situation as "a shambles", while Baker admits: "Of course we are very concerned about the future".

Despite the uncertainty, a "special working group" of 20 scientists from the four partners is pressing on with two tasks assigned to them by the ITER council. The first, which is more or less complete, determined that a smaller version of ITER — known as 'ITER-Lite' — costing \$5.5 billion instead of \$10 billion, could meet many of the project's technical objectives (see *Nature* **393**, 406; 1998).

The second task, requiring the group to explore other collaborative experiments short of that, is proving more difficult to execute. At Vienna, the United States pushed for the group to formally spell out such options. But ITER advocates in Europe and Japan fear that such a step will undermine the case for any version of ITER.

For many of its supporters, ITER's symbolism as an example of international collaboration in science is at least as important as its technical objectives. It is on the basis of that broader significance that the energy department is now trying to save the agreement in the Congress.

President Bill Clinton will not, however, veto the energy and water appropriations bill in order to save the agreement. And its fate is likely to have little impact on support for the US magnetic fusion research programme, which will be funded next year at close to this year's level of \$229 million. **Colin Macilwain**

## Fall in Australian R&D is linked to tax law change

[SYDNEY] After 11 years of sustained growth, business expenditure on research and development has fallen sharply in Australia, according to two independent surveys by business and government.

The fall is being linked to the coalition government's cut in a tax concession for research and development in industry, from 150 to 125 per cent, two years ago. At the time, the government claimed it was needed to correct alleged — though never proved — abuses of the system through 'creative accounting' by some claimants.

An analysis by the Business Council of Australia (BCA) concludes that the concession was a "super-efficient vehicle for encouraging business expenditure on research and development, as business decision-makers tend to overestimate the benefits to their company's shareholders".

The BCA calculates that more than A\$1.5 billion (US\$908 million) in research and development has been lost since 1996 by 150 businesses with a total turnover of A\$125 billion; this is about one-third of the expenditure predicted if the concession had remained at the higher level. A survey by the Australian Bureau of Statistics agrees with the BCA, claiming that business investment in research declined by 7 to 8 per cent in real terms in the first year after the cut (1996–97).

The tax incentive was introduced in

1985 by a Labor government, at a rate competitive with Asian nations such as Singapore and Malaysia, to boost Australia's low level of industrial research and development. The move was followed by an immediate increase in business investment in research.

Over the five years to 1995, business investment grew by 13 per cent a year in real terms, rising from 0.5 per cent to 0.8 per cent of gross domestic product. An industry spokesman for the opposition Labor Party, Simon Crean, says the government will have cut \$2 billion in incentives for research and development over the four years to 2000.

He accuses prime minister John Howard of leaving Australian industry "severely exposed in an increasingly competitive international environment". Australia's federal science minister, John Moore, has announced amendments aimed at "streamlining" the tax concession and has agreed to the BCA's call for a "summit" on business research and development next year.

The BCA reports that the hardest hit areas are in "the more strategic and speculative research and development". It says that nearly a third of business investment was in "research-intensive sectors, including pharmaceutical and biomedical manufacturing, electrical goods manufacturing and telecommunications". **Peter Pockley**

## Japan picks prominent physicist to lead education ministry

[TOKYO] Physicist Akito Arima, a former president of Tokyo University and an influential voice in recent debates on how Japan should manage its science, was last week appointed education minister in the cabinet of Keizo Obuchi, Japan's new prime minister.

His appointment as head of the Ministry of Education, Science, Sports and Culture (Monbusho) has raised hopes about the chances of much-needed reforms to the country's universities and changes in the way science is managed.

Arima's appointment was announced on 30 July, two weeks after he was elected to the Upper House of the Diet (Japan's parliament) as the top candidate of the proportional representation list of the ruling Liberal Democratic Party (LDP); this allocates seats to candidates according to the number of votes polled by the party.

Until May, 67-year-old Arima was president of the Institute of Physical and Chemical Research (RIKEN), the renowned



Arima: an outspoken supporter of science.

research institute overseen by the Science and Technology Agency (STA).

Arima has been a member of various government panels, including the councils for central education and administrative

reform. Given this background and the LDP's strong support for him during the election, his appointment is not surprising.

His supporters included the former prime minister Ryutaro Hashimoto, who resigned after his party's poor showing at the polls, and whose administrative reform plans Arima helped draw up last year. Hashimoto belongs to the same political faction in the LDP as Obuchi.

Arima has long been an outspoken supporter of Japanese science and a key

advocate of reforms in its organization. For example, he was responsible for introducing the first external reviews of a Japanese university, and helped to shape the 1996 Basic Law for Science and Technology which was designed to increase Japan's spending by 50 per cent by the year 2001.

His appointment is seen as increasing the chances of success of the impending merger between STA and Monbusho. Many researchers have been concerned about the merger, claiming that STA's 'top-down' approach and Monbusho's education-orientated 'bottom-up' approach to research are fundamentally incompatible (see *Nature* **390**, 327; 1998).

Scientists generally support Obuchi's choice of Arima, but many are concerned that the new cabinet may not last long. The latest public opinion poll conducted by *Asahi Shimbun* shows that there is only 32 per cent support for Obuchi, so Arima may lose the chance to make a significant impact during his term of office. **Asako Saegusa**



## India may retaliate over US expulsion of its scientists

[NEW DELHI] The Indian government has asked its research community to show restraint while it considers how to respond to the expulsion of seven Indian scientists from the United States.

Up to 75 more may be asked to leave under the new US policy of expelling Indian and Pakistani scientists working on US government-funded projects or in institutions if they are affiliated to their countries' nuclear organizations, even if the scientists themselves work in non-nuclear fields.

The US action, intended to punish India for having conducted nuclear tests, has drawn strong protests. India's science minister, Murli Manohar Joshi, said India "expected more mature behaviour from the US administration".

In a statement, he accused Washington of reducing science and technology "to a ball game for serving partisan interests", and hoped that the US and international scientific communities would make concerted efforts to end such "discrimination".

The expulsions also prompted condemnation from the upper house of the Indian parliament, with some members urging prime minister Atal Behari Vajpayee to retaliate by expelling US scientists in India. But India's science secretary, Valangiman Ramamurthi, says: "We do not want to take this battle to the street corner. We want to behave in a dignified and mature way."

Nonetheless, India's science ministry is looking at options to hit back at the United States. Possible targets include cancelling an agreement with the US space agency NASA to share data obtained from its Insat geostationary satellites (see *Nature* 391, 109; 1998). Also in jeopardy is an agreement with India's Department of Biotechnology under which US companies were hoping to test new male contraceptive drugs.

A US State Department official describes the prospect of retaliation as "regrettable", but adds that it will not affect US government policy. The seven expelled scientists, he says, are all affiliated to Indian atomic energy- and defence-related research organizations. They were working on research projects at the National Institute of Standards and Technology in Maryland.

They have been asked to leave by the end of the month. Two have already left, however, and one is planning legal action against the US university which he says gave him 24 hours to leave the country with his family.

"The way Indian scientists are being treated as bonded labour is an insult to the US scientific community and its academic freedom," says Ramamurthi. **K. S. Jayaraman**

## Xenotransplant experts face good and bad news

[LONDON] An international panel of scientists is due to meet in London today (6 August) to discuss the results of organ-transplant studies. These are likely to send mixed signals on whether it is safe to transplant animal organs into humans.

The meeting is a closed one-day workshop organized by the UK Xenotransplantation Interim Regulatory Authority (UKXIRA). The topics for discussion are expected to include "promising results" from an industry-funded study of 150 patients already treated with living pig tissue, such as islet cells for the treatment of diabetes.

But academic researchers from Britain are expected to announce the results of a study showing that a variant of a pig retrovirus previously thought to be safe can infect human cells under laboratory conditions.

The meeting comes a week after the British government released its guidelines on regulating xenotransplants. These say that the Secretary of State for Health will decide on each application for a human trial of an organ transplant, based on advice from UKXIRA. The authority, in turn, will be supported by a number of expert assessors.

The welfare of animals bred for human transplants will be included in UKXIRA's terms of reference. And the government's advisory committee on dangerous pathogens is to set up a working group to investigate the risks of infection from different forms of xenotransplant therapy. The guidelines endorse a Home Office ban on the use of primates as organ donors.

The move has been welcomed by companies developing animal organs for human transplants, as it allows them to apply to proceed with human trials. A statement from Imutran, the xenotransplant arm of the life sciences company Novartis, described the announcement as "an important step".

Today's meeting was arranged to discuss the latest research on whether pig viruses can transfer into humans. This is one of the main concerns holding up regulatory approval for animal-to-human transplants.

Researchers have been able to eliminate most non-inherited viruses from human transplant organs bred in pigs. But there are four known variants of a pig virus — the porcine endogenous retrovirus (PoERV) — found in the chromosomes of pigs that cannot be eliminated.

Two of the four variants of PoERV are known to be capable of crossing the species barrier and infecting human cells. But today's meeting is expected to hear of a third virus, not found in all pigs, that has been



found to infect a human cell line under laboratory conditions.

Better news for proponents of xenotransplants is expected from a study organized by Imutran of 150 patients who have been treated with living pig tissue. Samples were collected from patients worldwide. Each was analysed twice — by a laboratory in the United Kingdom and by one in the United States — to see if the pig retrovirus had transferred to any patient.

"The results are very interesting," says David Onions, an adviser to Imutran and director of the Glasgow-based laboratory Q1 Biotech, which analysed one set of samples. Onions says he cannot provide further details until the data are published. But he describes them as "encouraging".

Imutran is relying heavily on the results of this study for its plans to develop xenotransplant organs. The next step, according to a company statement, is to attach transgenic livers outside the body of a small number of patients waiting for a liver transplant, enabling the pig livers to work along the lines of a liver dialysis machine.

The pig liver would work for about 48 hours, allowing doctors extra time to find a suitable human liver for transplantation. "This study will then be reviewed to assess the safety and efficacy of the transgenic organs," says the statement.

Imutran is not the first to use such a technique. An analysis by the Institute of Cancer Research in London, of two cases of pig kidneys similarly attached to patients, has been accepted for publication in a medical journal.

Meanwhile, the US Food and Drug Administration is due to publish guidelines on xenotransplantation this autumn. These are expected to advocate long-term monitoring of recipients and a possible ban on the use of non-human primates (see *Nature Medicine* 4, 876; 1998).

**Ehsan Masood**

# Italy pulls plug on unproven cancer 'cure'

[MUNICH] The Italian Ministry of Health, forced by public pressure to conduct clinical trials on the controversial 'Di Bella' cancer treatment, announced last week that patients with the types of cancer tested in the trials will no longer be reimbursed for such treatment.

This follows the news that the trials have shown the therapy to be of no clinical value. None of the 134 patients in four clinical protocols showed any indication of remission during three months of treatment, and three-quarters of them have died.

But support for the therapy is still being voiced from many quarters, most strongly from right-wing politicians hoping to benefit politically from its popular backing.

The therapy, designed by physician Luigi Di Bella, is a cocktail of natural products and the expensive drug somatostatin, sometimes also with a low dose of the chemotherapeutic agent cyclophosphamide. Di Bella, now in his mid-80s, claims to have cured thousands of patients in the past couple of decades (see *Nature* 391, 217; 1998).

Controversy arose when his many followers began to demand that the government pay for the treatment, which costs up to US\$6,500 per month. The health ministry

initially refused even to conduct clinical trials with the therapy — a prerequisite for reimbursable drugs — because of the lack of a scientific basis for its effectiveness.

When the strength of public support for Di Bella forced the ministry's hand earlier this year, minister Rosy Bindi approved ten tightly designed multicentre trials in different types of cancer.

Di Bella and his supporters, known as Di Bellists, have refused to accept that the therapy does not work, despite the fact that another trial conducted by the Di Bella-friendly government of Lombardia, whose results were announced a few weeks ago, had proved similarly negative.

Di Bella formally approved the protocols of the government-sponsored trials before they began. But after the results of the first four protocols had been published — the remaining six will be completed by October — he said he had not read the documents before he signed them. He also insists that the protocols do not correspond exactly to his therapy, and that the trials are therefore not valid.

The local judge from Lecce in Puglia, who first ordered that the state pay for a patient to be treated with Di Bella therapy, has set up

his own inquiry into the validity of the trials, and has ordered Giuseppe Benegiano, director of the Higher Institute of Public Health in Rome, which is coordinating the trials, to testify this week.

Benegiano, who has been inundated with calls from people claiming that the trials were designed only to "count the dead", has published a request for those cancer sufferers who claim to have been cured to come forward for assessment, "so that we can also count the living".

The issue is receiving close attention in top political circles. Prime minister Romano Prodi has publicly welcomed the "clarity" of the results of the trials, while his political opponents in right-wing parties such as Forza Italia, hoping to gain populist support, have accused the centre-left government of wanting to discredit Di Bella to avoid having to pay for the treatment.

Benegiano says he hopes that public opinion, which has been staunchly behind Di Bella, will be swayed by the mounting evidence — including trials conducted by Di Bellists themselves — that the treatment does not work. "But it will take some time, because the Di Bellists seem incapable of taking no for an answer," he says. **Alison Abbott**

## Germany faces physics graduate shortage as students turn away

[MUNICH] The number of new physics graduates in Germany is expected to fall by almost two-thirds in the next five to six years, according to the German Conference of Physics Faculties (KFP).

Konrad Kleinknecht, a spokesman for the KFP, warns that the decreased output will be well below the demands of industry and research. Gerhard Soff, vice-dean of the physical faculty of the University of Dresden, says university research could be in danger, as financing is linked to student numbers.

The news closely follows a similar fall in enrolment in chemistry courses (see *Nature* 388, 707; 1997), although general student numbers will have fallen by only 10 per cent.

Concern over employment prospects and the German phenomenon of *Technologie-feindlichkeit* (aversion to technology) are the most widely cited reasons for this.

The closing of the former East Germany's Academy of Sciences in 1990 threw many physicists and chemists out of work, against a background of world recession and loss of business for companies supplying military contractors after the end of the cold war.

Recently the situation has changed. "Graduates are practically snatched away from universities at the moment," says Kleinknecht. Industry needs young physicists in traditional areas such as

semiconductor technology and optics, as well as in new fields such as information technology and software design, he says. Management consulting and insurance companies are employing growing numbers of graduates. But, he adds, scepticism about new technologies is keeping many talented young people from choosing physics.

Bernd Fischer, a personnel officer at Siemens, reflects a widespread feeling among physics professors and in industry that school teaching is excessively influenced by

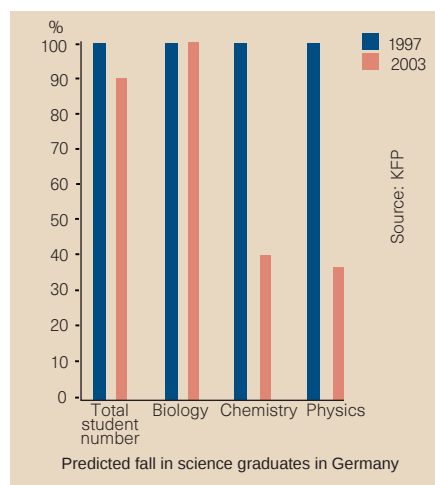
"the 1968 generation of teachers", who are not interested in science.

Ten scientific societies in Germany have formally requested science ministries in the 16 *Länder* (states) to take steps to increase and improve the teaching of mathematics and natural sciences at high schools, and to ensure that more pro-science teachers are employed.

The KFP is optimistic that campaigns can influence public opinion. The fact that biology has not lost popularity is at least partly ascribed to an energetic government campaign to promote biotechnology.

But Josef Langer, general secretary of the German University Rectors Conference, points out that disproportionate falls in student numbers in physics or chemistry are not unique to Germany. "The trend has already been observed in other northern European countries," he says. These include Austria, Belgium and the Netherlands.

A recent internal European Commission report notes the low status of science in some countries, but says the causes of falling enrolments are complex and "need a much deeper socio-cultural study". It says that education ministries, teachers and the media should work to improve the image of the natural sciences by emphasizing their positive effects on society. **Annette Kloboucek**



Figures show exact sciences falling from favour.

# Pesticide tests on humans cause concern

[WASHINGTON] The US Environmental Protection Agency (EPA) said last week that it was "deeply concerned" that some pesticide manufacturers appeared to be engaging in health-effects studies on human subjects "as a way to avoid more protective results from animal tests".

The statement came as it emerged that a Californian pesticide company has helped to pay for a study at the University of California, Davis, of human volunteers who were exposed to methyl isothiocyanate, the active ingredient in metam sodium, a potent soil fumigant (see box).

The company, Amvac Chemical Corporation, belonged to a coalition of six chemical companies — including Zeneca Ag Products — which spent roughly half a million dollars on the experiment, designed to meet California regulations. The study was carried out in 1994 and involved 70 paid volunteers.

Revelations last week that the same company had funded human experiments on British volunteers of an organophosphate pesticide, dichlorvos, have already focused attention in both Washington and London on a previously little-asked question: should pesticide makers be allowed to carry out tests on human volunteers to provide data to federal regulators, and, if so, under what conditions?

In response to news of the US-funded tests in Britain, the EPA said in a statement that "no human test data have been used by EPA for any final decisions about acceptable levels of pesticide use" under new food safety laws that critics claim the companies are trying to circumvent through human testing. (California's approach appears less stringent).

According to Ken Cook, president of the Washington-based Environmental Working Group (EWG), which advocates tougher pesticide controls, the Californian study is the first known pesticide experiment to have been conducted on human volunteers in the United States in recent years. He describes the experiment as further evidence that the EPA "needs to turn on its radar system and find out the extent of human experimentation on pesticides".

The British tests were described in a report by EWG that documents the testing on humans of dichlorvos for Amvac by the Medeval Laboratories in Manchester in 1997. *The Guardian* newspaper reported this week that human tests were still being carried out on the organophosphate insecticide azinphos-methyl by the Inveresk Clinical Laboratory in Scotland, for which students and others were being paid \$480 (US\$780).

In its report, EWG criticizes the EPA for using data from the British experiments in determining pesticide safety levels, and calls for a moratorium on the use of human data in pesticide regulation. "EPA is accepting and

evaluating human experimental studies that it does not require and, in fact, actively discourages," the report says.

The EWG report contends that EPA used the human data in deciding to relax the safety margins for aldicarb. The group also alleges that industry is deliberately using human data to circumvent the tougher safety standards required in the Food Quality Protection Act of 1996, which overhauled pesticide regulation. By testing on humans, it says, companies hope to avoid an extra, tenfold margin of safety in determining safe exposure levels.

The EWG report also criticizes the EPA for failing to ensure that pesticide companies abide by government ethics rules — the so-called Common Rule for human subjects' protection — adopted by the agency in 1991.

Amvac defends the use of humans in experiments by arguing that it meets the conditions in the Helsinki Declaration for Testing and Protection of Human Subjects, requiring informed consent of participants and approval by independent ethics boards.

Ian Chart, the company's director of regulatory affairs, says that testing pesticides on humans "is good science when carried out under the Declaration of Helsinki", which makes no distinction between the testing of pharmaceuticals and pesticides.

The company points out that dichlorvos, the pesticide used in the Manchester experiment, has been used for about 30 years to treat schistosomiasis.

Meredith Wadman

## California laboratory backs its use of volunteers

[WASHINGTON] In tests conducted at the University of California in Davis for Amvac, Zeneca and other chemical companies (see above), investigators sought to determine the levels at which methyl isothiocyanate (MITC), the active ingredient in the pesticide metam sodium, can be smelt, and those at which it produces physiological effects.

Eye irritation was measured by running varying concentrations of MITC vapour into goggles worn by volunteers for between one minute and eight hours. Tearing, blinking, subjective discomfort and redness were measured. The study found "no significant increases" in redness or tearing for most test subjects; irritation was deduced from blink rates and subjective discomfort.

The lead investigator, Michael Russell, a neurobiologist at the university's medical school, says the study — which was filed with the state government in September 1996 — produced "no ill effects whatsoever". Russell, who himself volunteered, says the study was approved without problem by the institutional review board at the medical school. He estimates that volunteers were paid \$150 or \$300,

depending on how long they spent wearing the goggles.

MITC — a less potent cousin of methyl isocyanate, the agent in the Bhopal disaster — is a toxic fumigant that kills virtually all soil organisms. Farmers use it to strip the soil of weeds, insects and microbes before planting. In humans, at high concentrations, it is very irritating to the eyes, skin and respiratory system.

In recent years, dozens of California residents living near farms using metam sodium have been hospitalized with sore throats, burning eyes, breathing difficulties, headaches, nausea and vomiting. In 1991, a train accident dumped at least 13,000 gallons of the pesticide in the Sacramento river, killing tens of thousands of fish. When tests showed that high doses of metam sodium cause birth defects in rats and rabbits, there were calls to pull it from the market.

The companies funded the Davis study because they were dissatisfied with animal data — in particular, a study of tearing in cats — that were being used by California's Department of Pesticide Regulation (DPR) for risk assessments. Russell says that the companies' interest was "to have an [approved exposure] level that is as

high as possible but still actually safe".

Veda Federighi, a DPR spokeswoman, stresses that the department "did not request a human subjects study" from the companies. But she adds, "we would have no problem accepting a [human] study that's done within appropriate guidelines". Federighi says the California DPR applies a tenfold safety factor to animal data to arrive at acceptable human exposure levels. "Often [companies] feel that that is inappropriate."

Critics say that metam sodium is such a potent poison that the government should be banning it, not using human data to regulate it. "We ought not to be exposing people to try to keep its use legalized," says Ralph Lightstone, a staff attorney with the California Rural Legal Assistance Foundation, an advocacy group for farm-workers.

California's DPR placed restrictions on its use in 1994, but the amount used grew in 1995, the most recent year for which data are available. Zeneca Ag Products, based in Wilmington, Delaware, stopped producing metam sodium recently; Amvac continues to make it, along with dichlorvos (see main story).

M. W.



## Confirmation of US science posts lets Colwell in at NSF



[WASHINGTON] The US Senate has confirmed appointments to several key science policy positions in the Clinton administration, ending fears that nominations would be held up until September or even later.

Neal Lane (above), former director of the National Science Foundation (NSF), was confirmed unanimously on 31 July as director of the White House Office of Science and Technology Policy (OSTP). Rosina Bierbaum was confirmed as OSTP associate director for environment. Bill Richardson, US ambassador to the United Nations and former New Mexico congressman, was confirmed as energy secretary.

Lane's confirmation will allow Rita Colwell, a marine biologist from the University of Maryland, to take over as the first woman and the first biologist to head the NSF. The confirmations took place on the last day before the Senate left town for its August recess.

## Gore pledges \$38 million for smoking research

[WASHINGTON] The US vice-president Al Gore has announced that \$38 million will be spent over the next two years on research at the National Cancer Institute into smoking prevention and cessation. In a speech last week to a conference on nicotine addiction at the National Institutes of Health, Gore excoriated the tobacco industry for "consciously lying" for decades about nicotine's addictive nature.

The research funds — including money to study the genetics of nicotine addiction — represent "not just a policy priority, but... a moral obligation", Gore told the conference, co-sponsored by the National Institute on Drug Abuse and the Robert Wood Johnson Foundation. Thomas Lauria, a spokesman for the Tobacco Institute, a trade group, called Gore's accusations "specious... political bluster" devoid of content.

## At long last: molecular pharmacology clarified

[MUNICH] To help bring some scientific order to the mushrooming field of molecular pharmacology, the International Union of Pharmacologists last week launched its definitive *Compendium of Receptor Characterization and Classification* at its thirteenth international congress in Munich.

Its aim is to describe the characteristics of cloned hormone and neurotransmitter receptors at a time when more data are being generated than ever before and terminology is inconsistent and confusing. Each receptor in the compendium is allocated a receptor code, analogous to the enzyme codes introduced in the 1950s by the International Union of Biochemistry and Molecular Biology. The receptor codes convey both structural and operational information.

## 'Too much secrecy' at Dounreay, say MPs

[LONDON] An all-party British parliamentary committee last week supported the government's decision in April to receive a consignment of mostly fresh nuclear fuel from Georgia at the nuclear reprocessing plant at Dounreay in Scotland. But members of the House of Commons Trade and Industry select committee concluded in their report that 'there remains too great a culture of secrecy' at and around the facility. Britain agreed to accept the material following concerns about its safety in Georgia. However, the decision was not made public until after the nuclear fuel had arrived in Scotland (see *Nature* 392, 850; 1998).

## Russia to pay up money owed to scientists

[MOSCOW] The Russian government has promised to pay all debts to the Russian Academy of Sciences (RAS) this month. The agreement follows negotiations with trade unions from RAS. The finance ministry will transfer 1 billion roubles (\$160 million) to RAS accounts: 800 million roubles for salaries and the rest to cover debts for the second quarter of 1998. Viktor Kalinushkin, an RAS trade-union leader, says scientists may join striking miners, teachers and nuclear workers this autumn "if the cabinet fails to fulfil this agreement".

## Animal guidelines annoy Indian researchers

[NEW DELHI] Proposed guidelines on the use of animals in research and a proposal to ban the import of laboratory animals are causing resentment among researchers in India. Under the guidelines, no institute will be able to acquire or use animals in research without the permission of a government regulatory committee. Animal experiments for teaching and training are also banned.

The guidelines are understood to be the work of Maneka Gandhi, the welfare minister, who is an ardent animal activist. Gandhi also chairs the Committee for the Purpose of Control and Supervision of Experiments on Animals.

Sandip Basu, director of the National

Institute of Immunology in New Delhi, describes the proposed rules as "ridiculous". Another leading scientist fears this will shut down animal experimentation.

## Britain delays controls on sale of vitamin B6

[LONDON] The British government is to delay plans to restrict the sale of vitamin B6 by at least 18 months. The decision follows last month's report by the House of Commons agriculture select committee, which said the plans were based on flawed science.

A new expert group on vitamins and minerals has been charged with investigating the safety of the vitamin. Plans to restrict sales were based on advice from the government's advisory committee on toxic chemicals in food, which had based its advice on research that claimed that high doses caused side effects such as numbness (see *Nature* 389, 7; 1997).

The agriculture select committee concluded that the committee on toxic chemicals seemed unable to distinguish between good and bad science.

## Brussels seeks 25 per cent cut in CO<sub>2</sub> emissions

[LONDON] The European Commission wants Europe's automobile industry to make a 25 per cent reduction in carbon dioxide emissions from all new cars within a decade. The requirement, if backed by European Union member states, will commit car makers to reducing carbon dioxide emissions from 185 grammes per kilometre in 1995 to 140 grammes per kilometre in 2008.

The 15 member states have agreed collectively to reduce greenhouse gas emissions by 8 per cent by 2010.

## Greenpeace 'tried to buy an atom bomb'

[MUNICH] The environmentalist group Greenpeace tried to buy an atomic bomb for \$250,000 from a Russian army officer in 1991, according to a former US military intelligence officer who joined the group in 1989. Breaking years of silence, William Arkin said last week that Greenpeace wanted to publicize the ease with which nuclear weapons in the former Soviet Union could fall into the hands of despots or terrorists.

Arkin says that a 29-year-old Russian officer had claimed that he could steal a warhead from an arsenal in East Germany and deliver it to Greenpeace before escaping to Sweden. But he vanished without trace before the deal came off.

"Since the whole action failed at such an early stage we felt no need to inform the public," says Norbert Schnorbach, a spokesman for Greenpeace Germany.



# Finding the funding for agriculture

*Sir* — I appreciate the timely appearance of the article and editorial on the support of US agricultural science (*Nature* 394, 207 & 210–211; 1998). They should serve to inform the scientific public about how science at the US Department of Agriculture (USDA) is currently funded and the uncertain future of the authorized new monies that would provide additional funding for plant and animal genome research (among other topics). The article also stressed the beleaguered state of the National Research Initiative (NRI), the department's competitive grants arm, where I am chief scientist on leave from the University of Missouri.

I want to emphasize that the NRI does more than support plant science. There are 29 panels, covering all aspects of agricultural research, from economics to genomics. Yes, the plant people have a

tough time compared with the typical investigator funded through the US National Institutes of Health (NIH), but, with ingenuity, they can get support from the National Science Foundation or even the Department of Energy. There is still some plant science funded by the NIH, if the work is properly dressed up to be health related.

There are, however, certain areas where the NRI is about the only source of federal funds. Consider, for example, the individual working on an agricultural project to study the interaction between grazing, plants, soils and water, or university scientists studying animal diseases that are not 'models' useful in human medicine. There are also broad areas of animal science that are understandably unpalatable to NIH; ruminant nutrition or lactation don't

exactly appeal to study sections, and who is going to fund competitive grants for genomics of agriculturally important animal species, if not the NRI?

I am on temporary (60%) assignment to the USDA. I can afford to do this only because of the indulgence of my university and because I am well funded by NIH on projects that, in a more logical setting, would be of at least equal interest to agriculture as to human health.

A sorry footnote is that the NRI may lose a further \$5 million of its already small appropriation because of an amendment to pay for a fiscal obligation elsewhere in the USDA unrelated to competitive grants.

**R. Michael Roberts**

158 Animal Sciences Research Center,  
University of Missouri, Columbia,  
Missouri 65211, USA  
e-mail: robertsrm@missouri.edu

## 101 uses for a natural history museum

*Sir* — The Briefing about natural history museums was very informative but the editorial, "101 uses for a dead bird", promotes a number of misconceptions (*Nature* 394, 105 & 115–119; 1998).

Museum collections are, as you recognize, the ultimate basis of most if not all biodiversity studies, including those that permit the decline in biodiversity to be monitored and controlled. Principal input is through the discipline of systematics, that is, the recognition of species, their variation and their relationships. Systematics condenses the vast amounts of disparate information in large numbers of specimens from many localities and puts it in comprehensible and usable form. Museums also act as repositories for other data including those on species distribution.

Systematics and related data collection have been pursued with increasing pace in natural history museums since their foundation, often in the nineteenth or even eighteenth centuries. Strong internal cultures within museums have helped ensure continued systematic output, even in conditions of considerable adversity. Any implication that museums and similar institutions have not previously been conducting what are essentially biodiversity studies is wrong; they are the very places where such investigations have been predominantly carried out.

You also state that museums have been parochial and "can no longer afford to work

in isolation". In the vast majority of cases, they never were and never did. There has been extensive global cooperation between large museums for nearly 200 years. For example, George Boulenger, the lone curator of lower vertebrates at the Natural History Museum in London, exchanged information in ten years around the turn of the century with some 700 correspondents, many in museums in other countries, including 20 in Italy alone. Such large-scale interaction is far more extensive today. Not only information but study specimens constantly pass between museums around the world and have done so for many years. In the Natural History Museum, many hundreds of loans involving tens of thousands of specimens are dispatched annually.

There is already a great deal of cooperation between relatively rich and poor nations, including the case you mention of Mexico's National Commission of Knowledge and Use of Biodiversity, which sent scientists to museums worldwide to assemble data on Mexican specimens. The commission is to be congratulated on producing cheaply a biodiversity management system by making use of these data, but it must be remembered that the information was readily available because museums cooperated willingly and actively in the project. More importantly, they had previously put a far greater amount of effort and expense into accumulating and curating the material, painstakingly identifying it and subjecting it to further systematic analysis, resulting in the description of many new species, and

making the data easily accessible. After all this, putting the information into a database as the Mexicans did was comparatively easy.

**Nicholas Arnold**

Department of Zoology, Natural History Museum,  
Cromwell Road, London SW7 5BD, UK

## German ambivalence to genetic engineering

*Sir* — You reported the results of a German project on perceptions of biotechnology, saying that Germans are deeply suspicious of genetic engineering and that expectations are predominantly negative (*Nature* 393, 299; 1998). But our study shows that the predominant attitude of Germans towards genetic engineering is in fact ambivalence.

Half the population thinks that genetic engineering is neither good nor bad. Along with this ambivalent perception of genetic engineering in general, there are rather clear, and differentiated, reactions to specific applications of genetic engineering. Fewer than 2% of the sample either agree with or reject all applications.

Medical and pharmaceutical applications are accepted at the same level as other positively evaluated, uncontested technologies, for example computer and information technologies. Agricultural applications are still not accepted.

**Jürgen Hampel**

Center of Technology Assessment in Baden-Württemberg, Industriestrasse 5,  
D-70565 Stuttgart, Germany

# Vision in the eternal present

Robert Ward

**When we are searching our visual environment, it might be expected that we accumulate records or memories of the locations that have been scanned. But a new study indicates that visual search can operate without the guidance of such memories.**

As quickly as you can, decide whether or not Fig. 1 contains the letter T. For decades<sup>1</sup>, it has been thought that visual-search tasks such as this proceed in much the same way as a well-organized person might search a pile of puzzle pieces for the one that fits a particular hole. A piece is first selected and compared to the target. If the piece fits, the search is over; if not, the piece is laid to one side so that it won't be selected again, and a new piece is selected. By putting the pieces aside, the searcher can remember which ones have already been examined. But a less organized or more easily frustrated person might simply throw the pieces back into the pile if they don't fit. This search could go on indefinitely, returning time and again to pieces that have already been assessed. Such an 'amnesic' approach to finding something appears to be strikingly inefficient. Yet, as reported by Horowitz and Wolfe<sup>2</sup> on page 575 of this issue, the visual system seems to operate in just this way.

The method used by Horowitz and Wolfe is simple and clever. In difficult visual-search tasks, such as that shown in Fig. 1, the time taken to find a target increases more or less linearly with the total number of items in the display. The speed of perceptual processing, in terms of average time per display item, is measured by the slope of the line that relates the search time to the number of items (the 'search function'). Horowitz and Wolfe measured this slope under two conditions. In the 'static' presentation, all items remained in the same location — as is typical in search experiments. But in the 'random' presentation, the target and all of the other display items changed position roughly every 100 ms.

If search performance depends on remembering which locations have already been examined, the continual changes in the random presentation should be disastrous. The effects of this disruption will be relatively small with few items in the display, and get increasingly worse as more items are added. So, the standard, memory-based model of search predicts that there will be much greater search slopes with random presentations than with static ones. But, in three experiments, Horowitz and Wolfe found equivalent search slopes for both types of presentation. In other words, an amnesic model of search seems

to be correct. As the authors describe it, the visual system as exposed in these tasks appears to operate in an 'eternal present'.

The results pose a puzzle. In these tasks, the visual system operates with no regard for or memory of previous visual states, even when this could be beneficial. Yet visual and memory systems are clearly not entirely separate — we can form memories of what we see, and can guide visual processing based on remembered locations and forms. So how can we explain the dissociation? Perhaps the visual system is designed to take nothing for granted. In the real world, where previously unoccupied locations may suddenly contain a bear or parking-meter attendant, it could be a mistake to rely too heavily on memories from previous visual analysis.

Not only can assumptions based on memories of a changing world be potentially misleading, but there is also a computational cost to constructing and maintaining these memories. Even if a visual stimulus is static and unchanging, the corresponding images projected on the retina are certainly not, as the observer moves their eyes, head and body. Loosely speaking, two strategies are available to deal with this changing series of images. One is to continually compute transforms that allow the stimulation received within a retina-centred frame of reference to be mapped onto a set of coordinates centred

on some stable external point of reference<sup>3</sup>. The amnesic strategy might be to forego this computation and get updates, as needed, from the environment itself.

For some tasks the amnesic approach may be sufficient. Consider Herbert. He's a robot whose sole function is to roam through a crowded office building and collect empty soft-drink cans<sup>4</sup>. Herbert has what might be called a profoundly amnesic visual system. If Herbert senses a can, he moves towards it; if he senses an obstacle, he moves around it. Herbert does not maintain memories of where cans or obstacles can be found, and would have no basis for realizing that a can or obstacle had moved or even changed form. Nevertheless, Herbert is reasonably effective at his simple job, responding to unforeseeable changes in the environment without the computational costs of memory.

Of course, human behaviour — and the demands on our visual system — is much more varied than Herbert's. Unlike Herbert, we may have basic, opposing tensions on the need to recognize change. On the one side there is the necessity to respond to the present visual state, regardless of what might have occurred in previous states. A system that operates solely in the present is well suited to meet this challenge. On the other side, change in itself often indicates where and how behaviour should be directed. It may be prudent to investigate the sudden appearance of new objects. Further research may show how the visual system balances the pressures to live in the present while remembering what is useful about the past. □

Robert Ward is at the Centre for Perception, Attention and Motor Sciences, School of Psychology, University of Wales at Bangor, Bangor, Gwynedd LL57 2DG, UK.

e-mail: r.ward@bangor.ac.uk

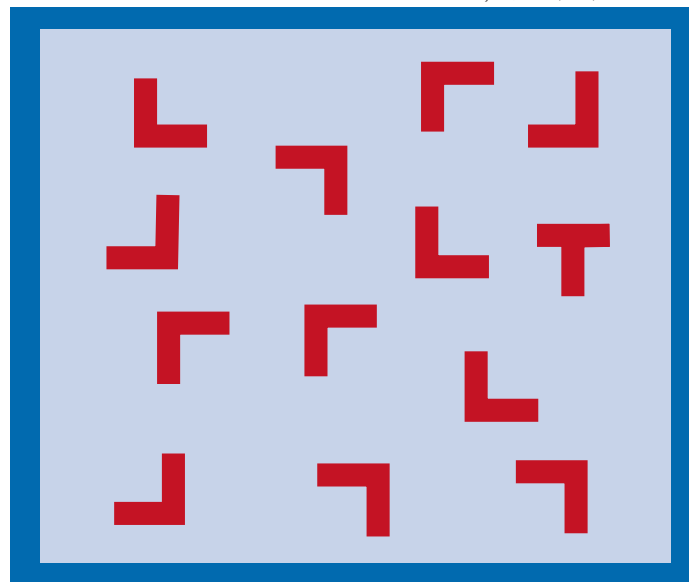
1. Sternberg, S. *Acta Psychol.* **30**, 276–315 (1969).

2. Horowitz, T. S. & Wolfe, J. M. *Nature* **394**, 575–577 (1998).

3. Andersen, R. A. *Cerebral Cortex* **5**, 457–469 (1995).

4. Brooks, R. A. *Robotics Auton. Sys.* **6**, 3–15 (1990).

Figure 1 Visual-search task. Horowitz and Wolfe<sup>2</sup> suggest that, contrary to expectations, in searching this figure for the letter T, your visual system does not remember which positions in the figure have already been scanned.



# NATURE

## 100 YEARS AGO

A remarkably fine specimen of the gigantic centipede (*Scolopendra gigas*) may now be seen in the Zoological Society's Insect House. It is not, perhaps, quite full grown, but measures about eight inches in length. It is fed principally on small mice, which it devours with alacrity. This specimen was captured in Trinidad, and forwarded to the society by Mr. R. R. Mole, Port-o-Spain.

\*\*\*\*\*

The expedition sent out to the Galapagos Islands, at the suggestion of the Hon. Walter Rothschild, last year brought home a fine series of living tortoises, which have been recently deposited in the Zoological Society's Gardens. There are in all fifty-two specimens belonging to the group of large land tortoises, namely thirty-three of *Testudo vicina* from the south part of Albermarle Island, and nineteen of *Testudo ephippium* from Duncan Island. These have been placed in the old Tortoise House in the North Garden, and feed greedily on cabbages. From *Nature* 4 August 1898.

## 50 YEARS AGO

There is a legend that [Gustav] Fechner was a physicist who was afflicted in middle life by an illness which affected his sight and probably his reason, and he thereafter gave up science and took to religious mysticism. Dr. Lowrie's narrative explodes the legend. Fechner's best-known religious work, "Life after Death", was published four years before his illness, and work on Fechner's law was done several years after. All his life he found the mechanistic view of physical science intellectually stifling, and fought against it; first with parodies and jokes, afterwards more seriously and constructively. He took a pan-physic or pan-theistic view of the natural world rather characteristic of the Romantics of his period. Unlike most of them, however, he did not turn his back on science; instead he tried to enlarge the scientific insight. Fechner was rash in speculation and perhaps entirely mistaken in thinking that scientific thought could transcend its traditional limits. The mistake was not that of a man who is just stupid ... Physiologists should note the strange diet of raw lean ham soaked in wine and lemon juice which initiated Fechner's recovery from his digestive troubles. From *Nature* 7 August 1948.

## Planetary science

# Io writes its history in hot metal

Lionel Wilson

The flow of data from the Galileo spacecraft continues to challenge our thinking about Jupiter's satellites. The latest news comes from Io: infrared data analysed by McEwen *et al.*<sup>1</sup> in *Science* show that the lava flows on Io are far hotter than all previous estimates, whether taken from the earlier Voyager missions<sup>2</sup>, contemporary telescopic observations<sup>3</sup> or earlier Galileo measurements<sup>4</sup>. Temperatures in the range 1,700 to 2,000 K or more are inferred, similar to those of the primitive, magnesium-rich komatiite magmas that came from partial melting of the mantle during Earth's early history. These hot magmas tell us something about Io's structure, and its history.

Io's intense volcanic activity is driven by tidal heating. Jupiter's gravity raises a tidal bulge in Io's body, and as the satellite rocks back and forth the bulge distorts its interior, and viscous dissipation generates the heat. Without the other satellites, Io would just settle into a circular orbit in which it keeps one face fixed on Jupiter, and there would be no heating; the vital rocking motion is maintained by a gravitational resonance between Io, Europa and Ganymede. It isn't clear, however, how long this resonant situation has existed, and we don't know the exact location or the consequences of heating inside Io.

It may be that heating has gone on for so long that all of Io's outer layers have been remelted many times over<sup>5</sup>, building up a 50-km-thick crust of light, highly evolved rocks, similar to the silica-rich rhyolite lavas and granite intrusions found in continental areas on Earth. In support of this, some spectroscopic data<sup>6</sup> appear to be consistent with the presence of silica- and alkali-rich rocks in regions far from currently active volcanic vents. But rocks with these compositions melt easily, and erupt at around 1,000 K, several hundred degrees lower than the temperatures now observed. Also, such lavas are very viscous on Earth and extrude slowly to form relatively thick flows that do not extend for more than a few kilometres. These characteristics are at odds with those of the lava flows now seen on Io's surface, whose thicknesses are clearly closer to 10 m than 100 m, and which commonly extend for up to 200 km from their vents — both factors implying a low lava viscosity and high eruption rate<sup>7,8</sup>.

So in this picture (where long-term extended heating produces an evolved crust) the lavas now being observed would have to be derived by melting a mantle enriched in magnesium, iron and calcium, underlying the silica-rich crust, in order for them to have the temperatures measured. There are great problems in persuading dense, metal-rich magmas to rise through a lower-density



Figure 1 Io, against the background of Jupiter's clouds. This false-colour image from the Galileo spacecraft shows sites of volcanic activity as dark spots, some marked by red deposits from explosive eruptions. Most of the dark material is in lava flows erupted at high temperature.

crust, but it commonly occurred in the early history of the Moon. To do this, high pressures must be generated in the magmas by the density decrease on melting<sup>9</sup>, and melting must be sufficiently fast for there to be no time for the high pressure to be released by plastic deformation in the surrounding rocks.

But McEwen *et al.*<sup>1</sup> prefer an alternative picture for Io. The high-temperature lavas might come from a mantle that has not undergone much differentiation, other than the separation of an iron core very early in Io's history. This could imply that tidal heating has not been so intense or continuous, although there is no other evidence to support this idea. If little differentiation has taken place, the crust would be more metal-rich and therefore denser, and the dense magma would have less trouble erupting. This idea is supported by the admittedly sparse six-colour spectroscopic data from Galileo's Solid State Imaging device, which can be interpreted to imply the presence of the magnesium-rich mineral pyroxene in dark deposits from the high-temperature vents. A very high percentage of partial melting of the mantle — about 40% — would be needed to produce lavas with the observed compositions and temperatures, but this is

NASA



exactly what happened during the Earth's early history to produce the very fluid lavas called komatiites.

A corollary of this idea is that Io's deeper mantle must be well mixed with the rocks at shallower depths, which implies efficient recycling of the surface layers into the interior. On Earth, partial crustal recycling is accomplished by plate tectonics, which involves the subduction of ocean-floor crust at the margins of continents. But no hint is seen on Io of the topographic features that would be expected if something akin to plate tectonics is taking place there. Instead, some other process, perhaps simply very rapid burial under new eruption deposits, must be at work<sup>10</sup>.

Whatever the recycling mechanism, there must be a link between the degree of partial melting of the mantle, the range of depths over which it occurs, and the rate at which the molten rock is extracted from

the solid residue and brought to the surface in new eruptions. It is this aspect of the origin of volcanic melts that is still least well understood on Earth, so the comparison of eruptions here and on Io should improve our understanding of volcanism in general. □

Lionel Wilson is in the Environmental Science Department, Lancaster University, Lancaster LA1 4YQ, UK.  
e-mail: l.wilson@lancaster.ac.uk

1. McEwen, A. S. et al. *Science* **281**, 87–90 (1998).
2. Pearl, J. C. & Sinton, W. M. in *Satellites of Jupiter* (ed. Morrison, D.) 724–755 (Univ. Arizona Press, Tucson, AZ, 1982).
3. Spencer, J. R. et al. *Geophys. Res. Lett.* **24**, 2451–2454 (1997).
4. McEwen, A. S. et al. *Icarus* (in the press).
5. Keszthelyi, L. & McEwen, A. S. *Icarus* **130**, 437–448 (1997).
6. Carlson, R. W. et al. *Geophys. Res. Lett.* **24**, 2479–2482 (1997).
7. Blaney, D. L., Johnson, T. V., Matson, D. L. & Veeder, G. J. *Icarus* **113**, 220–225 (1995).
8. Davies, A. G. *Icarus* **124**, 45–61 (1996).
9. Head, J. W. & Wilson, L. *Geochim. Cosmochim. Acta* **56**, 2155–2175 (1992).
10. Carr, M. H. et al. *Icarus* (in the press).

## Evolutionary biology

# Green beard as death warrant

Alan Grafen

The Gradgrindian view of science — facts come first, ideas later — is rarely true, and there are few more direct demonstrations of that than the discovery of a phenomenon that previously existed only in theory. Genetics is rich in such phenomena, and an example is provided on page 573 of this issue<sup>1</sup>, where Keller and Ross describe how they have identified evidence of a 'green-beard' gene. Quibbles over how one defines these genes make it hard to say whether this a 'first', and Keller and Ross themselves cite another candidate<sup>2</sup>.

The idea of green-beard genes was invented by Hamilton<sup>3</sup>, and named by Dawkins<sup>4</sup>. A basic question about the genetics of social actions is how a gene can spread by recognizing and interacting with individuals carrying copies of itself. The likely way is through kinship, and inclusive fitness theory contains many empirical examples of this. But, to isolate the phenomenon in its starkest form, Hamilton proposed a gene that does three things. It causes its bearer to display an observable and distinctive trait (the 'green beard'); to distinguish between individuals that do not display the trait and those that do; and to be altruistic to those that do. This gene has the capacity to recognize individuals with copies of itself, and helps them, and so helps itself.

This thinking belongs firmly in the 'selfish gene' tradition. It makes many geneticists feel uncomfortable, mutter grimly to each other that these so-called genes are completely hypothetical and wonder why anyone should bother to discuss a gene when it does not really exist. The work of Keller and Ross

brings these two worlds into collision.

They have studied the red fire ant, *Solenopsis invicta*, about whose population genetics, behaviour and ecology much is now known. This native South American species is of interest practically as well as scientifically, for it is a recently introduced pest in North America. (I can testify to the pain caused by the pheromonally synchronized bites of workers swarming out of an upturned canoe in a swamp in Georgia.) The ant has many strange features, among them being the existence of two subpopulations that are genetically connected only through males<sup>5</sup>. One subpopulation is monogyne (a single queen per colony), the other is poly-



Figure 1 Red fire ant queen and workers — here as attendants, not executioners. (Photo courtesy W. Tschinkel.)

gyne (several queens per colony).

Keller and Ross<sup>1</sup> report only on the polygyne ants, and show that the locus *Gp-9*, which has alleles *B* and *b*, is always heterozygous, *Bb*, in those queens that live long enough to lay eggs. Queens and workers that are *bb* seem to die young from physiological causes, but the remarkable findings concern how *BB* queens meet their fate. They are executed by workers after their emergence from the pupa, and Keller and Ross's experiments suggest that *Bb* workers are particularly incriminated in the deed. Whether the locus responsible is actually *Gp-9* or a very tightly linked locus is not yet known.

The connection with green beards is this. Through the mediation of the bearer, a gene (*b* or a tightly linked allele) induces killing of individuals that do not contain it (*BB* queens) while not inducing killing of individuals that do (*Bb* queens). The green beard in this case seems to be a chemical carried on the queens' cuticle, as Keller and Ross found that rubbing a worker against a *BB* queen causes it to be attacked by other workers, whereas rubbing it against a *Bb* queen does not. This is not exactly what is expected from the classical green-beard principle, in which green-beard wearers would kill non-wearers. Nonetheless, the mechanism of advantage to the gene is impressively similar.

One puzzle is why such a gene should be found at all. Hamilton and Dawkins both emphasized that their theoretical point did not rely on such an unlikely gene actually existing. Biologists didn't used to suppose that organisms had the capacity to detect other individuals' genes and respond accordingly. For one gene to create such a system and manipulate it in a particular way does indeed seem far-fetched, but this capacity no longer surprises us. Genetic discrimination can underlie kin recognition<sup>6</sup>, species recognition and individual recognition, and has parallels in immunology. Once such a capacity exists in a species, a mutant that tweaks the existing system could plausibly have the properties of a green-beard gene. In his work, Haig<sup>2</sup> has proposed the existence of green beards in the context of genetic conflicts between parents and their offspring during mammalian fetal development. Now that he and Keller and Ross have stimulated us to reflect on the matter, it is not so surprising that a green-beard gene should be found.

What kind of functional system might the *Gp-9* locus be involved in? Queens in the monogyne subpopulation are all homozygous *BB*, and live to lay eggs, so the locus may be closely involved in maintaining monogyny and polygyny, and genetic separation of subpopulations. It is intriguing to speculate that *BB* queens, if allowed to live, would attempt to make the nest monogynous by attacking the other queens or inciting the workers to do so — this is, after all, presumably how monogyny is maintained in monogynous nests.

Moreover, in the experiment each *BB* queen remained alive while kept in a small colony fragment with a few hundred workers, and was executed only when she was returned to the main colony with its existing queens. This supports the idea that workers in a queenless colony would tolerate a *BB* queen. If this is right, it is possible that genotype-specific execution is a trait evolved by workers to permit polygyny in some nests, in a species in which the monogyny-enforcing mechanism is still active in other nests.

Keller and Ross describe their green-beard gene as an 'outlaw'. Admittedly, the comment is only made in passing, but are they correct? In this context an outlaw is usually defined as a gene whose action favours itself, but opposes the reproductive interests of the individual organism. Where there are outlaws, natural selection at different loci is pulling the organism in different directions. Theoretically speaking, green-beard genes

are not, in general, outlaws<sup>7</sup>. Do the effects of *b* harm the interests of the workers or reproductive queens? Does the definition need elaboration for eusocial species, when some notion of colony interest might be invoked?

The interactions between theory and reality are multi-layered. Keller and Ross's study reminds us that scientists who, like Haldane, feel sad when beautiful hypotheses are slain by ugly facts, can draw comfort when beautiful ideas step boldly out into the living world. □

Alan Grafen is at St John's College, University of Oxford, Oxford OX1 3JP, UK.  
e-mail: alan.grafen@sjc.ox.ac.uk

1. Keller, L. & Ross, K. G. *Nature* **394**, 573–575 (1998).
2. Haig, D. *Proc. Natl Acad. Sci. USA* **93**, 6547–6551 (1996).
3. Hamilton, W. D. J. *Theor. Biol.* **7**, 17–52 (1964).
4. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, New York, 1976).
5. Ross, K. G. *Genetics* **145**, 961–974 (1997).
6. Grosberg, R. K. & Quinn, J. F. *Nature* **322**, 456–459 (1986).
7. Ridley, M. & Grafen, A. *Anim. Behav.* **29**, 954–955 (1981).

## Earth science

# Hot stuff under southern Chile

Julie D. Morris

The isotopes of the uranium decay series are a wonderful tool for exploring geological processes that have taken place on the Earth's surface and in its interior<sup>1</sup>. On page 566 of this issue<sup>2</sup>, Sigmarsson and colleagues describe how they have taken advantage of these isotopes to 'image' the melting of a slab of oceanic plate being subducted along a stretch of southwestern South America. The results tell us how melting of the slab relates to volcanism at the surface, and may also bear on our knowledge of the generation of continental crust in the distant past.

The modern history of studies of subduction began in the 1960s with plate tectonics, which showed that chains of active volcanoes around the Pacific rim are underlain by subducting oceanic lithosphere. Earthquakes track the ocean plate, formed at mid-ocean ridges, as it travels beneath the arc of a volcanic chain and into the deep mantle (see Fig. 1). The ubiquitous association of active volcanic arcs and subducting slabs led quite naturally to the belief that lavas erupted at the Earth's surface were formed by melting of the subducting slab about 110 km below the volcanic chain. Hence the consternation when, in the early 1970s, it emerged that most erupting lavas have the wrong chemistry to be the result of slab melting<sup>3</sup>.

Since then, a more complex model for the relationship between subduction and modern arc volcanism has been developed. While moving from mid-ocean ridge to subduction trench, the ocean plate incorporates sea water. Water is bound in hydrous minerals in the basaltic layer and in hydrous sedi-

ments deposited atop the plate, and also occupies the space between sediment grains. As the oceanic plate (which is usually old and cold) subducts, the water between sediment grains is squeezed out. The hydrous minerals formed at low pressure and temperature transform to minerals that are stable at increasing pressure and temperature, or

break down and release their water to the overlying plate. The oceanic plate itself is generally too cold to melt beneath the arc.

The behaviour of water during the subduction cycle may affect a variety of events (the localization of chemosynthetic biota near the trench; the distribution of hazardous inter-plate earthquakes between 0 and about 40 km depth; the explosivity of arc volcanoes; and the formation of ore deposits). Most notably in this context, however, water-rich supercritical fluids or sediment melts released from the slab lower the melting point of the mantle wedge above the slab<sup>4</sup>, with the wedge then being the source of the lavas that add to the mass of the continents today. Strikingly, the composition of continental crust forming now does not appear to match the average composition of the crust formed in the past, suggesting that processes today may be different from those that operated when the bulk of the continental crust was formed.

Just occasionally, rocks in active volcanic arcs have a chemical signature which indicates that they were formed by melting of the subducted oceanic plate itself. Such rocks are called 'adakites'<sup>5</sup>, and their chemical fingerprint<sup>6,7</sup> shows that modern ones have many features in common with ancient continental crust. This fingerprint is taken to reflect melting of hydrous basalt or its metamorphosed equivalent — although whether melting occurs at the base of the arc crust, or deeper, at the top of the subducting slab, is a matter of debate. Often, adakites are reported from margins where the subducting slab is

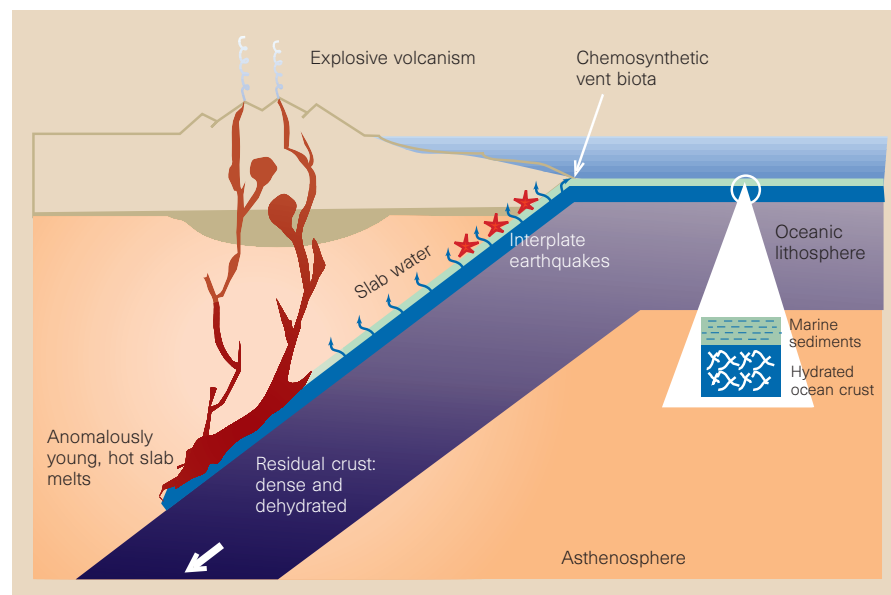


Figure 1 Slab melting beneath modern volcanic arcs. Many textbooks include diagrams such as this, but the scheme applies only to a few modern arcs, where very young and hot oceanic crust is subducting and melting at depth. The austral Andean volcanic zone studied by Sigmarsson *et al.*<sup>2</sup> is one such setting, and the magmas erupting here stem from the slab itself and have chemical compositions similar to ancient continental crust. At most convergent plate margins, supercritical water-rich fluids move from the slab to the mantle wedge, where they initiate mantle melting. Note that marine sediments often subduct with the plate, and that the slab is subducted past the arc to the deep mantle.

## Galactic structure

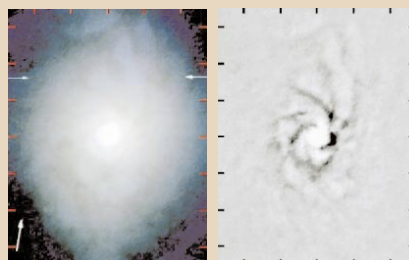
## The music of the spirals

These may be pictures of sound waves, say B. G. Elmegreen *et al.* in the 20 August issue of *Astrophysical Journal Letters* (503, L119–L122; 1998).

Both images show the inner 3,000 light years of the spiral galaxy NGC2207. The one on the right has been processed to highlight the dark dust lanes, which form a tangled spiral pattern. Nothing odd about finding spirals in a spiral galaxy, you might think — but these are different.

There is no sign of rapid star formation, which occurs in the huge spiral arms that define such galaxies. And the conventional spiral-arm-making mechanism — in which correlated stellar orbits produce spiral density waves, compressing interstellar gas and so triggering the formation of bright new stars — doesn't work so near to the nucleus.

Instead, Elmegreen *et al.* propose that



these spirals are amplified sound waves. Gentle acoustic waves are generated further out in the galaxy, perhaps by supernova explosions going off in star-forming regions. As they travel inwards, they should be focused and amplified by the disk-like geometry, to become strong disturbances such as these. If so, they would constitute an impressively loud bass sound, about 56 octaves below middle C.

Stephen Battersby

young and slow-moving. Under these conditions, the top of the slab at depth is modelled as being hot enough to melt<sup>8</sup>, a condition expected to be more common early in Earth's history. In short, the rock-forming processes of modern adakites may shed light on ancient continental growth, when the Earth was hotter.

This is where Sigmarsson *et al.*<sup>2</sup> come in. They describe uranium-series disequilibria isotope data for the austral volcanic zone (AVZ) of the southernmost Andes, one of the rare localities in which young (Holocene) adakites occur<sup>9</sup>. They then compare data from the AVZ with data from the southern volcanic zone (SVZ), located north of the intersection between the Chile ridge and trench (see the map on page 567). The age of the subducting plate is similar in the two volcanic zones, but the plate is subducting much faster beneath the SVZ and no adakites are reported there.

In the uranium isotope series, the decay from uranium or thorium to lead occurs through a series of smaller decay steps, producing a number of intermediary parent–daughter pairs with geologically useful lifetimes of up to about a million years. The <sup>238</sup>U–<sup>230</sup>Th activity disequilibria that Sigmarsson *et al.* report for the AVZ are a feature that will disappear if more than 350,000 years have elapsed since the two isotopes fractionated.

The data<sup>2,10</sup> show several notable features. The SVZ lavas have a pronounced U excess, seen only (although not always) in the magmas above subduction zones. Such an excess is thought to indicate the transport of hexavalent U in preference to tetravalent Th in oxidizing water-rich fluids, moving from slab to mantle wedge, within the past 350,000

years. The AVZ lavas, by contrast, all show Th excess, which is unusual in arc lavas, but common in the basalts of mid-ocean ridges and ocean islands. In the latter settings, Th excess is attributed to melt equilibration with mantle at depths great enough to stabilize garnet (which partitions U preferentially to Th in its mineral structure<sup>11,12</sup>). The crust beneath the AVZ is estimated to be up to 35 km in depth<sup>13</sup>, so, if seismic methods ultimately agree, any basalt on the base of the crust is too shallow to stabilize garnet. Taken together, the Th excess, the general chemical characteristics of the AVZ lavas and the hot subducting slab all imply that the adakites here are generated by slab melting.

It remains to be seen if the simple systematics for the AVZ lavas hold when a larger sample set is analysed. If they do, we can surmise that the slab beneath more than 400 km of the southernmost Andes is melting to a uniform degree. During subduction, the basalt of the oceanic crust will metamorphose to amphibolite (rock that is rich in the hydrous mineral amphibole) or to eclogite (clinopyroxene and garnet rock). Amphibole is likely to be the main host for water in the subducting slab at depth in either case. The uniform Th excess observed in the AVZ lavas can be modelled by a constant 20% melting of the slab, if garnet is abundant in the residual slab and amphibole is completely consumed during the melting. Estimates of the precise pressure and temperature of slab melting depend on the actual bulk composition, mineralogy, and water content and distribution (free or bound in minerals) of the slab at depth — adakitic melts could be generated between about 60 and 96 km depth and between about 800 and 1,150 °C (refs 8, 14). Under these pressures and tem-

peratures, the AVZ slab would be several hundred degrees hotter than predicted for most subducting slabs<sup>8</sup>.

We are, of course, left with several open questions. Most importantly, will the simple chemical relationships reported here still apply when a larger data set is analysed? In passing, the authors note the geographically systematic chemical variations in the AVZ. If this is a result of slab melting, does the incoming plate show a geographical pattern of chemical variation that matches that seen in the volcanoes? And if so, why? Stern and Killian<sup>9</sup> attribute geographically controlled chemical variations in the AVZ in part to increasing interaction between slab melts and the overlying mantle as well as the crust of the upper plate. Can the U-series systematics truly remain unchanged during mantle and crustal assimilation? The patterns seen in the U-series data are a combination of U–Th fractionation as well as of the time that has elapsed since fractionation; how, in detail, do the two effects combine to generate and preserve a trend that is parallel to the isotope equiline depicted in Fig. 2 of the paper (page 568).

Nonetheless, some of Sigmarsson and colleagues' results are arresting. Seismic methods do not reveal the subducting slab at depth beneath the AVZ, because the slab is too hot to generate earthquakes big enough to record at a distance. A local seismic network may ultimately image the slab, but geochemical methods do it now — they show the slab at depth, reveal that it is melting, and constrain its temperature. The U-series data suggest that the deep slab is dehydrated and rich in the dense mineral garnet, which may affect the geodynamics of the plate subducted beyond the arc. And, as the authors point out, modern adakites such as these are the best opportunity to study the active processes that gave rise to continental crust earlier in Earth's history. □

Julie D. Morris is in the Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130-4899, USA.  
e-mail: jmmorris@levee.wustl.edu

1. Ivanovich, M. & Harmon, R. S. (eds) *Uranium-Series Disequilibrium* (Clarendon, Oxford, 1992).
2. Sigmarsson, O., Martin, H. & Knowles, J. *Nature* **394**, 566–569 (1998).
3. Gill, J. B. *Contrib. Mineral. Petrol.* **43**, 29–45 (1973).
4. Tatsumi, Y. *Geophys. Res. Lett.* **13**, 717–720 (1986).
5. Kay, R. W. J. *Volcanol. Geotherm. Res.* **4**, 117–132 (1978).
6. Defant, M. J. & Drummond, M. S. *Nature* **347**, 662–665 (1990).
7. Drummond, M. S. & Defant, M. J. *J. Geophys. Res.* **95**, 21503–21521 (1990).
8. Peacock, S. M., Rushmer, T. & Thompson, A. B. *Earth Planet. Sci. Lett.* **121**, 227–244 (1994).
9. Stern, C. R. & Killian, R. *Contrib. Mineral. Petrol.* **123**, 263–291 (1996).
10. Sigmarsson O., Condomines, M., Morris, J. D. & Harmon, R. S. *Nature* **346**, 163–165 (1990).
11. Beattie, P. *Nature* **363**, 63–65 (1993).
12. LaTourette, T. K., Kennedy, A. K. & Wasserburg, G. J. *Science* **261**, 739–742 (1993).
13. Kay, S. M., Ramos, V. A. & Marques, M. J. *J. Geol.* **101**, 703–714 (1993).
14. Rapp, R. P. & Watson, E. B. *J. Petrol.* **96**, 891–931 (1995).



## Innate immunity

## Plants just say NO to pathogens

Jeff Dangl

Every home gardener can attest to the voracity of plant pathogens, but less well-known are the specific mechanisms that plants deploy to resist such fungal, bacterial, viral and nematode infections. Chief among these is the plant 'innate immune system', which is largely based on the specific recognition of pathogen-encoded molecules through probable receptors encoded by disease-resistance (*R*) genes<sup>1</sup>. A battery of cellular responses is then engaged (Fig. 1), ultimately leading to death of the pathogen<sup>2</sup>. Papers by Delledonne *et al.*<sup>3</sup> (page 585 of this issue) and Durner *et al.*<sup>4</sup> (appearing soon in *Proceedings of the National Academy of Sciences*) now add to our understanding of the signal-transduction events that mediate these defence mechanisms and ensure continued production in the garden.

A plant's cellular responses to infection are often correlated with rapid, localized cell death — termed the hypersensitive response — at the site of attempted infection. This is a programmed cell death in the sense that it can be mimicked by plant mutants in the absence of pathogen<sup>5</sup>. But it is not clear whether the hypersensitive response itself kills the pathogen, or whether the response is a by-product of the execution mechanism.

One of the cellular responses to invasion is an oxidative burst, leading to the production of reactive oxygen intermediates (ROIs). We do not know whether the ROIs orchestrate the hypersensitive response directly (by killing cells<sup>6</sup>), indirectly (by signalling further cellular responses<sup>7</sup>), or both. Superoxide<sup>8</sup>, the first product of the oxidative burst, is a poor candidate for the cell executioner — it has a short half-life, being rapidly converted to hydrogen peroxide, and it does not readily diffuse. Hydrogen peroxide itself is toxic at high concentrations, but not enough of it is commonly produced during resistance responses to kill cells outright<sup>7,9</sup>.

Given these doubts about whether, at the concentrations generated during the oxidative burst, ROIs could kill cells outright, a co-conspirator has been sought. In mammalian macrophages, ROIs collaborate with nitric oxide (NO) to execute bacterial pathogens<sup>10</sup>. Delledonne *et al.*<sup>3</sup> and Durner *et al.*<sup>4</sup> now indict NO as the molecule that collaborates with ROIs to trigger transcriptional activation of plant defence genes and the hypersensitive response.

Durner and colleagues measured the activity of NO synthase in tobacco leaves infected with tobacco mosaic virus. They found that NO synthase is induced in resistant, but not susceptible, plants. Delledonne *et al.* extend this finding to cultured soybean

cells, measuring both the activity of NO synthase and the subsequent release of NO in response to either a bacterial pathogen or molecules that elicit defence responses. Thus, plants produce NO during *R*-gene-dependent resistance responses, probably using an enzymatic machinery similar to that found in animal cells.

Both groups show that production of NO is enough to activate parts of the defence response. When Durner *et al.* injected animal NO synthase, its substrate and co-factors into tobacco leaves, a common marker of defence-gene transcription called PR-1 was activated. This activation was mimicked by NO donors and blocked by NO scavengers or NO synthase inhibitors. Moreover, PR-1 was activated through accumulation of salicylic acid, an intermediate required for many disease-resistance responses (Fig. 1). Delledonne *et al.* found that NO acts synergistically with ROIs to increase death in soybean cells — ROIs generated exogenously or by a pathogen-driven oxidative burst require NO for maximal killing, and steady-state levels of only 1  $\mu$ M ROI are sufficient.

In the most enlightening experiments of the lot, Delledonne *et al.* show that NO synergizes not only with ROI, but also with an endogenous factor that is dependent on salicylic acid and potentiates the overall defence response<sup>11</sup>. This strengthens evidence that accumulation of salicylic acid can act as a

'gate' downstream of *R*-gene-dependent ROI production, and also as an 'amplifier' of post-ROI cellular responses. So, the collaboration between ROIs and NO establishes a binary switch for fine-tuning the magnitude of the signals that flow through the defence-response pathway. In this scheme, the hypersensitive response is the final manifestation of this amplification.

The ROI/NO collaboration in transcriptional activation of defence genes proceeds by at least two pathways<sup>4</sup>. For the induction of PR-1 at least, NO acts through salicylic acid. By contrast, NO can also drive the activation of phenylalanine ammonia lyase (PAL). Induction of PAL can lead to the synthesis of more salicylic acid, so, although PAL induction does not depend on salicylic acid, it could amplify the other processes that do. The NO-dependent induction of PAL is mediated by guanylate cyclase, acts through cyclic GMP, and can be further stimulated by cyclic ADP-ribose<sup>4</sup> (Fig. 1). These molecules mediate several NO responses in animals, and cGMP is critical for plant responses to light stress (and a variety of other cellular processes). Because cADP-ribose is involved in modulating calcium channels, it is perhaps no surprise that high levels of calcium can trigger NO synthase activity<sup>3</sup>. Hence, a downstream, NO-mediated rise in the concentration of calcium could also stimulate NO production.

How might this binary switch and runaway cycle of ROI/NO and salicylic acid lead to a hypersensitive response? One hypothesis builds on the finding that synthesis of NO from L-arginine can lead to the depletion of this amino acid. When levels of L-arginine go below a critical threshold, NO synthase

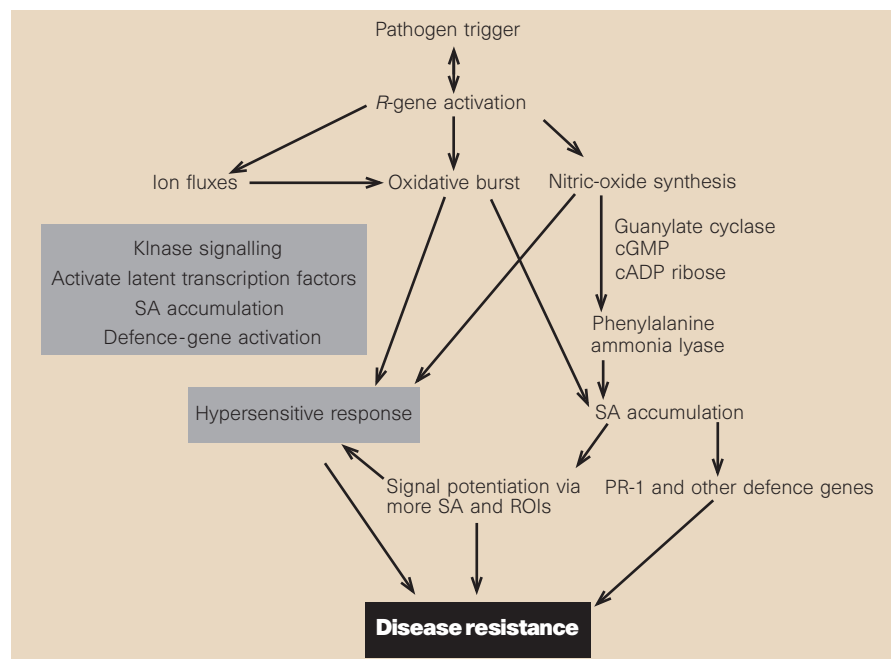


Figure 1 Plant defence responses to pathogen invasion. Defence responses<sup>14</sup> include: ion fluxes at the plasma membrane; an oxidative burst leading to production of reactive oxygen intermediates (ROIs); accumulation of salicylic acid (SA) in and around the infected cell; local transcriptional activation of defence genes; and hypersensitive cell death.

can also produce superoxide. The balanced synthesis of NO and superoxide allows the formation of peroxynitrite radicals, which are extremely potent radicals that can wreak substantial cellular damage<sup>12</sup>.

Delledonne *et al.* and Durner *et al.* have established NO as the concert master in the hypersensitive response and defence-gene activation. Yet several questions remain. First, to what extent is the generation of NO required to stop the pathogen's growth? Although Delledonne and colleagues provide tantalizing evidence that inhibition of NO accumulation might result in expanded pathogen growth, these data are the least convincing in the papers and will need to be extended to other pathogens and plants. Second, what parts of the genetically defined response pathway in the 'model' plant *Arabidopsis thaliana* are required for NO activation? Third, we still need to find NO synthase genes in plants, although the plant homologue of a protein that inhibits an animal NO synthase has been identified. Fourth, what actually kills plant cells during the hypersensitive response? And finally, just as some animal pathogens have learned to turn the toxic effects of NO into a virulence function<sup>13</sup>, necrotrophic plant pathogens have

probably evolved mechanisms for co-opting the ROI/NO collaboration to kill cells that they then can use for nutrients. The repartee of infection and response will continue to amuse and frustrate home gardeners and plant biologists alike. □

Jeff Dangel is in the Department of Biology and Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, CB#3280, Chapel Hill, North Carolina 27599-3280, USA. e-mail: dangel@email.unc.edu.

1. Bent, A. *Plant Cell* **8**, 1757–1771 (1996).
2. Hammond-Kosack, K. E. & Jones, J. D. G. *Plant Cell* **8**, 1773–1791 (1996).
3. Delledonne, M., Xia, Y., Dixon, R. A. & Lamb, C. *Nature* **394**, 585–588 (1998).
4. Durner, J., Wendehenne, D. & Klessig, D. F. *Proc. Natl Acad. Sci. USA* (in the press).
5. Dangel, J. L., Dietrich, R. A. & Richberg, M. H. *Plant Cell* **8**, 1793–1807 (1996).
6. Levine, A., Tenhaken, R., Dixon, R. & Lamb, C. J. *Cell* **79**, 583–593 (1994).
7. Jabs, T. *et al. Proc. Natl Acad. Sci. USA* **94**, 4800–4805 (1997).
8. Doke, N. *Physiol. Plant Pathol.* **23**, 345–357 (1983).
9. Glazener, J. A., Orlandi, E. W. & Baker, C. J. *Plant Physiol.* **110**, 759–763 (1996).
10. Nathan, C. *Cell* **82**, 873–876 (1995).
11. Shirasu, K. *et al. Plant Cell* **9**, 261–270 (1997).
12. Xia, Y. & Zweier, J. L. *Proc. Natl Acad. Sci. USA* **94**, 6954–6958 (1997).
13. Akaiki, T. *et al. Proc. Natl Acad. Sci. USA* **93**, 2448–2453 (1996).
14. Dangel, J. & Jones, J. D. G. (eds) *Curr. Opin. Plant Biol.* **1**, 274–365 (1998).

## Cancer

# Proteases – invasion and more

Dylan R. Edwards and Gillian Murphy

The June AACR/APMIS conference\* on 'proteases and protease inhibitors in cancer' provided an opportunity to take a good look at the importance of these proteins in cancer biology, particularly for functions other than cellular invasion<sup>1</sup>. The conference came at an appropriate time, with booming interest in angiogenesis (the formation of new blood vessels) as a therapeutic target and with first-generation synthetic matrix metalloproteinase inhibitors (MMPIs), such as British Biotech's marimastat, hitting the clinic. The emerging picture is that some questions — such as what individual proteases and inhibitors do during tumorigenesis, and whether this information is useful in prognosis or therapy — may soon be answered.

Proteases give cancers their defining characteristic — the ability of malignant cells to break out of tissue compartments. Most cancers originate from epithelial tissues in which oncogenes are activated or tumour-suppressor genes lost. Such genetically abnormal epithelium is supported by normal stromal cells (fibroblasts, inflammatory cells and blood-vessel cells) that become activated

during tumour progression to produce proteases, inhibitors and other regulatory factors (Z. Werb, UCSF). In fact, most of the increased protease production associated with malignant tumours comes from host stroma rather than the tumour cells themselves. But the cellular origins of the proteolytic machinery vary in different tumour types, leading Keld Danø (Finsen Lab., Copenhagen) to conclude: "Who has the weapons is not important — only that they are there".

Interactions between tumour and stromal cells control the two protease systems that are responsible for most of the proteolysis outside the cell. These are the urokinase plasminogen activator (uPA)/uPA receptor (uPAR)/plasminogen network, and the MMPs. Stromal uPA and MMP-2 are produced as inactive precursors and then activated on the surface of tumour cells (Fig. 1, overleaf), thus allowing malignant cells to breach basement membranes. These enzymes also facilitate the sprouting of blood vessels to feed the growing tumour (M. Pepper, Univ. Med. Center, Geneva). Anticancer therapies that target these stromal contributions to invasion, metastasis and angiogenesis attack a genetically stable cell population, so they may not encounter the resistance that is associated with the use of conventional chemotherapeutic drugs.

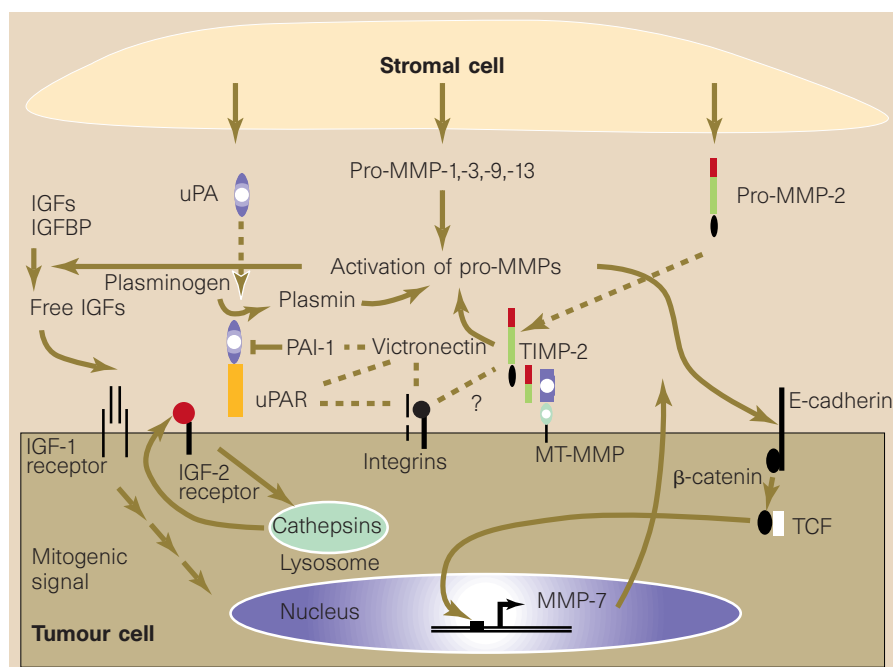
Does cancer invasion correspond to normal tissue remodelling? Very closely, if you consider comparable processes such as wound healing in the skin and squamous-cell skin carcinoma. Wound healing is delayed in mice that lack plasminogen, or in wild-type animals treated with a broad-spectrum MMPI (K. Danø). Moreover, healing is completely blocked when plasminogen-deficient mice are treated with MMPI, arguing that the uPA/uPAR/plasminogen and the MMP systems overlap functionally. This also means that the uPA/uPAR/plasminogen system can do more (in certain remodelling situations) than control the activation of MMPs, and this could be relevant to the treatment of certain types of cancer.

One component of the uPA/uPAR/plasminogen system — uPAR — has also been linked with chemotaxis, adhesion and signalling (F. Blasi, Univ. Milan). The uPA–uPAR interaction increases the binding of vitronectin to the cell, meaning that, somewhat paradoxically, uPA promotes cell adhesion<sup>2</sup>. To make matters even more complicated, PAI-1, which is the main PA antagonist, also binds to vitronectin and disrupts adhesion (D. Loskutoff, Scripps Inst., La Jolla). In this instance, PAI-1 acts as a cell-detachment factor, perhaps explaining why increased levels of PAI-1 correlate with a poor prognosis in cancer (J. Foekens, Rotterdam Cancer Inst.), and why PAI-1-deficient mice show reduced tumour invasion and vascularization compared with wild-type animals (J.-M. Foidart, Univ. Liège). So PAI-1 might be a switch that controls adhesion and migration through the uPA/uPAR system.

Another part of the plasminogen story is also one of the hottest new topics in cancer. The much-lauded angiogenesis inhibitor angiostatin<sup>3</sup> promotes apoptosis in endothelial cells, but not in other types of cell (M. Pepper). Angiostatin is a proteolytic breakdown product of plasminogen (through cleavage by plasminogen's active counterpart, plasmin, or certain MMPs). Curiously, mice that lack plasminogen show reduced rates of tumour invasion, growth and metastasis (T. Bugge, Children's Hosp. Res. Fdn, Cincinnati), although this probably reflects the malignancy-promoting role of plasmin. We also eagerly await the characterization of a putative angiostatin receptor (L. Holmgren, Karolinska Inst., Stockholm).

What have mouse models taught us about the involvement of MMPs in tumorigenesis *in vivo*? Overexpression of MMP-3 in the mammary gland leads to precocious development and increased tumour formation (M. Bissell, Lawrence Berkeley National Lab.). Conversely, rates of tumorigenesis are reduced in certain genetic backgrounds<sup>4</sup> by deletion of MMP-3. Overproduction of MMP-3 also leads to cleavage of E-cadherin and, thus, breakdown of interactions between epithelial cells<sup>5</sup>. As discussed by Lynn Matrisian (Van-

\*Proteases and Protease Inhibitors in Cancer, held by the American Association for Cancer Research and ACTA Pathologica Microbiologica Immunologica Scandinavica, Nyborg Strand Hotel, Nyborg, Denmark, 14–18 June 1998.



**Figure 1** Interactions between tumour and stromal cells, and the regulation of proteases and protease inhibitors. The two main protease systems are the urokinase plasminogen activator (uPA)/uPA receptor (uPAR)/plasminogen network, and the matrix metalloproteinases (MMPs). IGF, insulin-like growth factor; TCF, T-cell factor.

derbilt Univ., Nashville), loss of E-cadherin function allows transcriptional activation of MMP-7 (Fig. 1), which sets in motion further remodelling events. Mice that lack another MMP, MMP-9, show failures of vascularization and apoptosis in the skeletal growth plate early in development<sup>6</sup>. When these mice are crossed with animals in which expression of MMP-9 coincides with activation of an angiogenic switch during tumour formation, loss of MMP-9 suppresses tumorigenesis (Z. Werb). This is exciting evidence to link MMP-9 causally with angiogenesis.

The idea that some proteases mainly affect tumour growth rather than invasion may also apply to the cathepsins (H. Rochefort, INSERM, Montpellier). Cathepsins are aspartyl or cysteine proteases that are found within lysosomes in normal cells, but are released in high levels by many types of cancer cell. Cathepsins interact with the mannose 6-phosphate/insulin-like growth factor (IGF)-2 receptor, which is subsequently internalized. This prevents the receptor from acting as a sink for IGF-2, so the cell-proliferative actions of IGF-2 can proceed via the signalling IGF-1 receptor. The IGF signalling pathway may also be the target for the actions of MMPs and tissue inhibitors of metalloproteinases (TIMPs) on tumour growth. Overproduction of TIMP-1 suppresses T-antigen-induced hepatocellular carcinomas, possibly by preventing MMP-mediated degradation of IGF-binding proteins and, thus, reducing the local availability of IGF-2 (R. Khokha, Ontario Cancer Inst., Toronto).

What does all of this mean for cancer therapeutics? At least six synthetic MMPis

are now undergoing clinical trials, with phase I/II data published for marimastat (P. Brown, British Biotech, Oxford). This drug shows good safety and tolerability, although reversible musculo-skeletal toxicity is the main side-effect<sup>7</sup>. The preclinical data showing suppression of tumour invasion, metastasis and angiogenesis are impressive, and, with phase III trials continuing for at least another year, the problems at British Biotech seem to have more to do with management tactics than with marimastat itself.

Clearly, we need to know a lot more about the basic biology of proteases and inhibitors. By identifying the key 'weapons' at specific tumour sites, and their precise functions, the next generations of selective antagonists can be developed. The potential of combination anti-protease strategies using MMPis and promising synthetic uPAR antagonists (S. Rosenberg, Chiron Corp., Emeryville) needs to be studied, as does the integration of protease inhibition with conventional chemo- and radiotherapies. The era of protease and angiogenesis inhibitors in cancer therapy has only just begun. □

Dylan R. Edwards and Gillian Murphy are in the School of Biological Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK.  
e-mails: dylan.edwards@uea.ac.uk  
g.murphy@uea.ac.uk

- Chambers, A. F. & Matrisian, L. J. *Natl Cancer Inst.* **89**, 1260–1270 (1997).
- Chapman, H. A. *Curr. Opin. Cell Biol.* **9**, 714–724 (1997).
- O'Reilly, M. S. *et al.* *Cell* **79**, 315–328 (1994).
- Wilson, C. L. *et al.* *Proc. Natl Acad. Sci. USA* **94**, 1402–1407 (1997).
- Lochter, A. *et al.* *J. Cell Biol.* **139**, 1861–1872 (1997).
- Vu, T. H. *et al.* *Cell* **93**, 411–422 (1998).
- Nemunaitis, J. *et al.* *Clin. Cancer Res.* **4**, 1101–1109 (1998).

## Daedalus

### Smokeless powder

Smoking delivers nicotine to the smoker. Nicotine is pleasurable, which is why he smokes. However, smoking harms him, and annoys those around him. But nicotine is addictive, so he can't give up. This dismal reasoning underlies one of the fiercest public debates of our time. A new finding (M. L. Pianezza, E. M. Sellers & R. F. Tyndale *Nature* **393**, 750; 1998) offers a way out. Nicotine is largely oxidized in the liver to cotinine. People whose livers perform this reaction badly are much less likely to become addicted to smoking. Daedalus has three possible explanations of this finding, each with a useful twist.

If nicotine cannot go to cotinine in the body, it will go instead to nicotine *N*-oxide. Being very similar to nicotine, the *N*-oxide may be its antagonist — sitting on and blocking its receptors but failing to trigger their pleasure. This should protect the smoker from addiction. If so, it could form the basis of a splendid new anti-smoking treatment.

Alternatively, it may be that nicotine is the pleasurable agent of smoking, but it is cotinine that is addictive. Blocking the reaction to cotinine (by a suitable enzyme inhibitor) would then convert smoking from a degrading addiction to a voluntary pleasure. It would also prolong that pleasure. Deprived of its fastest degradation route, nicotine would last longer in the body. The smoker would be satisfied with fewer cigarettes.

Or cotinine might turn out to give both the pleasure and the addiction of smoking. This is the most revolutionary possibility of all. For smoking is a clumsy, annoying, inefficient ingestion process. It survives because nicotine is unstable in air; it cannot easily be packaged as a drink (like caffeine or alcohol) or a pill (like amphetamine). It must be thermally liberated from a stable precursor, and breathed in at once, before it degrades.

But cotinine is stable in air. So nicotine could be extracted from tobacco, converted to cotinine, and sold as a pill or a beverage. Smokers by the million would happily abandon their smelly habit, and satisfy their craving by a product that could be drunk or chewed. Their risk of lung cancer (triggered by all that tarry smoke) would plummet. So might their risk of heart disease, which is blamed on nicotine itself. Cotinine, a product of the body's detoxification mechanisms, should be far less cardiotoxic. The commercial possibilities are limitless. If Joe Camel doesn't do it, DREADCO will.

David Jones



# Shelley's shocks

**Electricity seemed new and magical, its limits not yet known: could it really arouse the dead? The Frankenstein story expressed an era's trepidation at the prospect of discovering the secret of life.**

**Martin Kemp**

The popular image of the magus-scientist who discovers the ultimate key to the creation of life has a long history, from the era of the alchemists to the more alarmist accounts of the recent cloning of Dolly. Never was there a greater sense that the secret of life was in the process of being disclosed than in the late eighteenth century, in the wake of revelations about the life-giving properties of 'oxygenated air' and sensational experiments with electricity.

*Frankenstein or The Modern Prometheus*, written by the 19-year-old Mary Shelley and published in 1818, is the supreme and most enduring literary product of these obsessions. During "long conversations" with Percy Shelley and Lord Byron in Byron's Swiss villa, she tells how "various philosophical doctrines were discussed, and among others the nature of the principle of life... They talked of the experiments of Dr [Erasmus] Darwin... who preserved a piece of Vermicelli... till by some extraordinary means it began to move with voluntary motion... Perhaps a corpse could be re-animated: galvanism had given a token of such things." The reference to galvanism is hardly surprising.

**Great interest centred upon Luigi Galvani's account in 1781 of how the detached leg of a frog could be made to move when an arc of two metals formed a bridge between the muscle and the crural nerve. Galvani's claim to have discovered animal electricity attracted an enthusiastic following. The chemical potency of electricity as it was being progressively disclosed seemed perfectly suited to be the mysterious ingredient that infused dead things with vital powers.**

A series of experimenters took up the challenge to produce life from death. The German physician, Karl August Weinhold, introduced a zinc and silver amalgam into the spinal chord of a kitten whose brain had been spooned out, with the result that it "hopped around, and then sank down exhausted".

Galvani's nephew, Giovanni Aldini, used electric currents to stimulate vivid expressions in the heads of executed criminals. And, in the year of the publication of *Frankenstein*, Andrew Ure, a Scottish chemist working in London, sensationally adapted Aldini's methods to induce 'life' in a criminal's corpse.

Shelley is careful not to describe Victor



Spark of life: animation as depicted in *Son of Frankenstein* (Rowland V. Lee, 1939) (left) and *Young Frankenstein* (Mel Brooks, 1974).



Frankenstein's "instruments of life", but it is clear that he used the unleashed powers of "electricity and galvanism" which Victor describes, after witnessing a bolt of lightning destroying an oak tree, as "new and astonishing to me". The repeated cinematic reworkings of the story typically characterize the mechanism that infused the "spark of life into the lifeless thing" as harnessing vast electrical power, sometimes directly from lightning.

What shocked Shelley, as she lay in bed imaging the terrible scene, still has the power

to haunt the public imagination: "I saw the hideous phantasm of a man stretched out, and then, on the workings of some powerful engine, show signs of life, and stir with an uneasy, half-vital motion. Frightful it must be; for supremely frightful would be the effect of any human endeavour to mock the stupendous mechanism of the Creator of the world." □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

# Triclosan targets lipid synthesis

Triclosan is a broad-spectrum antibacterial and antifungal agent<sup>1,2</sup>, which acts by previously undetermined mechanisms, that is used in products such as antiseptic soaps, toothpastes, fabrics and plastics. Here we show that triclosan blocks lipid synthesis in *Escherichia coli*, and that mutations in, or overexpression of, the gene *fabI* (which encodes enoyl reductase, involved in fatty acid synthesis) prevents this blockage. This is, to our knowledge, the first evidence that triclosan acts on a specific bacterial target, rather than as a nonspecific 'biocide'.

Five independent triclosan-resistant mutants of *E. coli* K12 strain AG100 (ref. 3) with different levels of resistance were isolated on LB agar plates containing 0.2 µg ml<sup>-1</sup> triclosan, a trichlorinated diphenyl ether (Irgasan, a gift from Ciba-Geigy) (Table 1). We prepared a genomic *Sau3AI* library from strain AGT11K (mutant AGT11 deleted for *acrAB*) in plasmid pBR322. Ten clones with inserts of various sizes all expressed the same level of triclosan resistance.

From the partial sequence of one clone, pLYT8, and from the *E. coli* genomic database, we identified two genes, *ycjD* and *fabI* (Fig. 1), in the insert. A deletion of *ycjD* (in clone pLYT12) or of most of *fabI* (in clone pLYT11) showed that triclosan resistance was associated with the intact *fabI* gene on pLYT12; the *tet* promoter from the vector was not required (Fig. 1).

The *fabI* gene encodes enoyl-acyl carrier protein reductase, an essential enzyme of relative molecular mass 27,900 that is involved in the synthesis of fatty acids<sup>4</sup>. Sequencing of the entire *fabI* gene of pLYT8 (residues 190–1,265 (numbering according to Bergler<sup>5</sup>), including the upstream 'BoxC' region) uncovered a single mutation, GGT to GTT, at codon 93 of the *fabI* gene. This mutation resulted in the conversion of glycine 93 to valine in the protein.

We replaced the mutation-carrying 606-base-pair *SspI*–*HindIII* fragment of mutant pLYT12 with its wild-type counterpart (derived from a KlenTaq polymerase (Clontech) polymerase chain reaction (PCR) product from parental strain AG100). The resulting sequence-confirmed plasmid, pLYT27, increased triclosan resistance in the host strain by 16- to 32-fold, showing that the wild-type gene has a clear multicopy effect, as would be expected if FabI were the target for the drug. As cells containing plasmid pLYT8 or pLYT12, each of which contains the Gly 93→Val mutation, were even more resistant (about 300-fold more resistant than wild type; Fig. 1), this mutation must indeed be responsible for the triclosan resistance in the original mutant AGT11.

**Table 1 Characteristics of triclosan-resistant *E. coli* mutants**

| Strain | <i>fabI</i> mutation | MIC (ratio to wild-type) |             |
|--------|----------------------|--------------------------|-------------|
|        |                      | Triclosan                | Diazaborine |
| AG100  | None (wild-type)     | 1.0                      | 1.0         |
| AGT8   | ND                   | 5.8                      | 7.1         |
| AGT9   | ND                   | 3.1                      | 3.6         |
| AGT11  | Gly 93→Val           | 95                       | 37          |
| AGT23  | Met 159→Thr          | 12.2                     | 0.6         |
| AGT25  | Phe 203→Leu          | 6.1                      | 4.3         |

Minimal inhibitory concentration (MIC) values, determined by broth dilution with twofold steps, are the average of three experiments and are expressed as ratios to the MIC for the wild-type strain AG100 (MIC for triclosan, 0.8 µg ml<sup>-1</sup>; diazaborine, 6 µg ml<sup>-1</sup>). ND, not determined.

The sequence of a *fabI* PCR product prepared from strain AGT23 (covering the same DNA region as was sequenced for pLYT8) revealed a single mutation (ATG became ACG), resulting in the replacement of methionine 159 by threonine in the FabI protein. Strain AGT25 also had a single mutation: phenylalanine 203 was replaced by leucine (TTC became CTC).

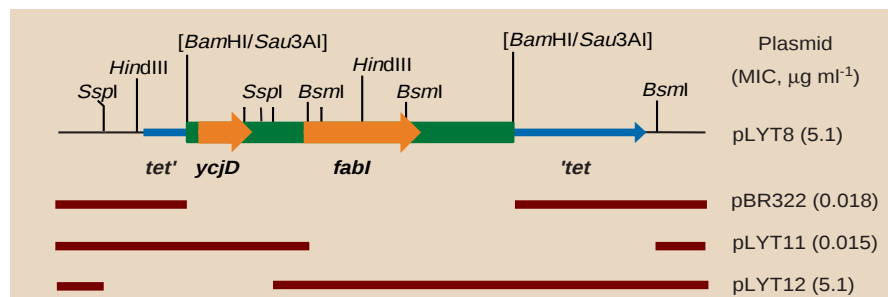
P1 transduction<sup>6</sup> of *zci-3118::Tn10kan* (ref. 7) together with wild-type *fabI*, both at position min 29 of the *E. coli* chromosome, into the two remaining (unsequenced) triclosan-resistant mutants, AGT8 and AGT9 (and into the sequenced mutant AGT11 as a control), led to loss of the triclosan-resistant phenotype at the same frequency for all of the mutants. Therefore, in the mutants whose DNA we have not yet sequenced, altered *fabI* is again probably responsible for triclosan resistance.

If FabI activity were inhibited by triclosan, fatty acid synthesis and consequently lipid synthesis should be reduced. We preincubated logarithmically growing cells of strain AG100 in LB broth with triclosan (or with other drugs as controls) for 8 minutes, and then added [1-<sup>14</sup>C]-acetate (sodium acetate, 59.5 mCi mmol<sup>-1</sup>, ICN Pharmaceuticals, added at 5 µCi ml<sup>-1</sup>) to assay lipid synthesis. Eight minutes later, we pipetted 0.08 ml of the culture onto a Whatman 3MM paper disc. We determined radioactivity levels after treating the discs in

cold 10% trichloroacetic acid (TCA) and washing them in TCA and 10% cold ethanol. All drugs were tested at concentrations just sufficient to stop the growth of AG100 an hour after their addition.

Ethanol (0.5%), present from the solvent used for most of the drugs, had little effect by itself. Diazaborine (see below), a specific inhibitor of FabI (8 µg ml<sup>-1</sup>; a gift from G. Hoegenauer), reduced acetate incorporation into lipids by 93%. Triclosan (0.24 µg ml<sup>-1</sup>) inhibited 92% of acetate incorporation, and the effect occurred within 3 minutes. In contrast, chloramphenicol (13 µg ml<sup>-1</sup>), a protein-synthesis inhibitor, and ciprofloxacin (0.045 µg ml<sup>-1</sup>), an inhibitor of DNA synthesis, reduced incorporation by only 19% and 2%, respectively. In the resistant mutant AGT11, lipid synthesis was blocked by only 2% with 0.24 µg ml<sup>-1</sup> triclosan and by 75% with 26 µg ml<sup>-1</sup> triclosan. These results are consistent with the target of triclosan being FabI.

Diazaborine inhibits FabI in *E. coli* and *Salmonella typhimurium*<sup>8</sup>; its binding depends on the presence of the co-factor NADH<sup>8,9</sup>. In a similar way, activated isoniazid inhibits InhA<sup>10</sup>, the *Mycobacterium tuberculosis* enoyl reductase, which shares significant sequence identity with *E. coli* FabI. A Gly 93→Ser mutation in *E. coli* FabI causes diazaborine resistance, reduces the binding of diazaborine to the enzyme<sup>8</sup>, and



**Figure 1** Restriction map (partial) and triclosan resistance of pLYT8 and its deletion mutants. In plasmid pLYT8 the thick (green and orange) horizontal region represents chromosomal DNA (*Sau3AI*-digested) inserted into the *Bam*HI site of the *tet* gene (blue) of the pBR322 vector. The deletion mutant pLYT11 was created using *Bsm*I, and deletion mutant pLYT12 was created using *Ssp*I. The minimal inhibitory concentration (MIC) of triclosan (in parentheses) was measured by agar dilution for each plasmid in hypersusceptible host AG100A (AG100  $\Delta$ *acrAB::kan*).



lowers the specific activity of the enzyme<sup>11</sup>. Our most triclosan-resistant mutant, AGT11, had a Gly 93→Val mutation. Its growth rate in LB broth was about 40% less than that of the wild-type parent, indicating that it may have a less active FabI enzyme.

The finding that mutations at residues 93, 159 and 203 lead to triclosan resistance correlates strikingly with the crystal structure of wild-type *E. coli* FabI protein<sup>12</sup>: all three of these residues line the cleft at which NADH binds. At this site, diazaborine is covalently linked to NAD<sup>+</sup> (ref. 12). Four of the five triclosan-resistant mutants were also resistant to diazaborine (Table 1), further indicating that FabI is the actual target of triclosan.

Triclosan lyses *E. coli*<sup>13</sup> and *Porphyromonas gingivalis*<sup>14</sup> cells. We found that such lysis needed higher drug concentrations than those needed to stop growth, judging by a 30–50% loss in the absorbance ( $A_{530}$ ) of susceptible cultures within two hours of the addition of triclosan, accompanied by a 4–5 log decrease in viability. Parental strain AG100 needed 0.15  $\mu\text{g ml}^{-1}$  triclosan to inhibit the growth rate by 50% but 8  $\mu\text{g ml}^{-1}$  for lysis, and the Gly 93→Val mutant AGT11 required 13  $\mu\text{g ml}^{-1}$  triclosan for 50% inhibition of the growth rate but did not lyse even when 256  $\mu\text{g ml}^{-1}$  was added; this amount far exceeds the solubility of the drug. These results indicate that the *fabI* mutation also offers protection against triclosan-mediated lysis.

Our results show that organisms that are intrinsically resistant to triclosan may contain triclosan-insensitive enoyl reductases. Like triclosan, other drugs that are at present thought to be nonspecific 'biocides' might actually have specific targets.

**Laura M. McMurtry, Margret Oethinger,**

**Stuart B. Levy**

Center for Adaptation Genetics and Drug Resistance, and

Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA  
e-mail: sleavy@opal.tufts.edu

12. Baldock, C. *et al. Science* **274**, 2107–2110 (1996).

13. Regos, J. & Hitz, H. R. *Zbl. Bakt. Hyg. I. Abt. Orig. A* **226**, 390–401 (1974).

14. Cummins, D. J. *Clin. Periodont.* **18**, 455–461 (1991).

## Oldest known fossils of monocotyledons

The monocotyledonous angiosperm clade (class Liliopsida) includes roughly 50,000 species<sup>1</sup> of diverse forms. The group comprises such economically noticeable plants as palms, orchids, most of the horticultural bulbs, and grasses, which include some of the most important food crops, such as maize, rice and other grains. Modern monocotyledons are diverse and dominate many habitats, but the fossil record of these plants is meagre, fossils of monocotyledonous flowers are rare, and the earliest putative monocotyledonous fossils (pollen and leaves) are all equivocal<sup>2,3</sup>. Here we describe the oldest known fossil flowers that can be definitely assigned to the Liliopsida.

Our understanding of the diversification of dicotyledonous angiosperms in the Late Cretaceous period has increased given fresh evidence of fossilized flowers from North America, Sweden and Portugal<sup>4–11</sup>. These fossils include taxa with affinities to several major clades of dicotyledons. Molecular and morphological analyses<sup>12,13</sup> suggest that the monocotyledon clade is an early offshoot from relatively primitive dicotyledons, and one might therefore expect to find evidence for concurrent monocotyledon and early dicotyledon diversification. However, monocotyledonous fossil flowers have not been reported previously from these mid and late Cretaceous deposits.

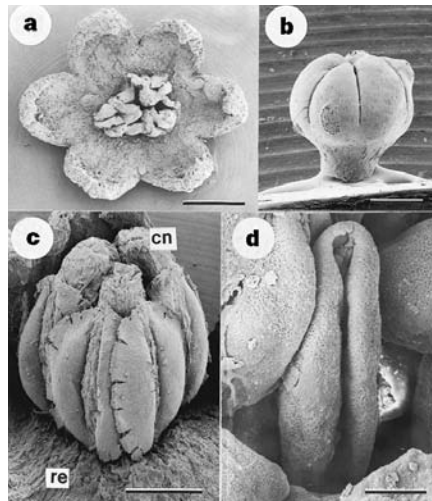
We have now recovered fossil flowers from fluvial sediments of the Turonian Raritan Formation, which is exposed in New Jersey, United States (Upper Cretaceous stage, approximately 90 million years before present)<sup>14</sup>. The nature of the associated palaeoflora (leaves, flowers and pollen) indicates that there was a subtropical to tropical palaeoclimate at the time of deposition. The monocotyledonous flowers are associated with abundant angiosperm dicotyledonous flowers and fruits representing at least 100 separate species (see, for example, refs 7–11), fern vegetative and reproductive structures<sup>15</sup>, and various leaves and reproductive structures of conifers<sup>16</sup>. The exquisite preservation of floral morphology and the often perfect retention of cell-by-cell anatomical details in these fossils allow us to study an array of characters and to compare them closely with those of modern flowers.

The fossil monocotyledonous flowers are minute, unisexual (only staminate flowers are known), trimerous, radially

symmetrical and pedicellate. They are 0.06–0.1 mm across and 0.04–0.08 mm high. The perianth comprises six basally fused tepals. The tepals are triangular in shape and valvate in bud; they are glabrous and lack stomata; and their apices curved toward their adaxial sides (Fig. 1a, b).

The androecium consists of three stamens arranged in the centre of the receptacle. The anthers are sessile, two-celled and extrorse, and they each dehisce to produce a longitudinal slit exposing a one-layered endothecium with U-type thickenings. The connective extensions are thick and well developed and extend beyond the anther sacs (Fig. 1c). Clusters of pollen grains are preserved within the open anther sacs. Pollen is boat-shaped (polar/equatorial diameter = 5  $\mu\text{m}$ /12  $\mu\text{m}$ ) and monosulcate with a fine granulose (Fig. 1d) and tectate exine, and the columellae are apparent in some broken grains. There is no indication of a gynoecium or pistillode(s).

The distinctive combination of features and organization of the floral organs indicates a close relationship of the fossils with members of the modern monocotyledon family Triuridaceae. We tested this relationship by including the new fossil taxon in a morphological cladistic analysis that includes 103 monocotyledon taxa (comprising families and genera), the fossil taxon, and 101 morphological characters<sup>13</sup>.



**Figure 1** Scanning electron micrographs of fossil flowers from the Upper Cretaceous stage of New Jersey. **a**, Top view of a mature and open flower. The tepals show the apex curved toward the centre of the flower. Note the central position of the androecium, which is formed by three stamens, in a single whorl. Original magnification  $\times 20$ ; scale bar, 600  $\mu\text{m}$ . **b**, Side view of a bud, showing the tepals basally fused and the stout pedicel. Original magnification  $\times 30$ ; scale bar, 300  $\mu\text{m}$ . **c**, Dissection of a bud; note the sunken anthers in the receptacle (re), extrorsely opened and with prolonged connective extensions (cn). Original magnification  $\times 50$ ; scale bar, 300  $\mu\text{m}$ . **d**, Monosulcate pollen found in one anther, showing the fine granulate exine. Original magnification  $\times 4,000$ ; scale bar, 3  $\mu\text{m}$ .

1. Furia, T.E. & Schenkel, A.G. *Soap Chem. Special* **44**, 47–50 and 116–122 (1968).
2. Bhargava, H. N. & Leonard, P. A. *Am. J. Infect. Control* **24**, 209–218 (1996).
3. George, A. M. & Levy, S. B. *J. Bacteriol.* **155**, 531–540 (1983).
4. Cronan, J. E. Jr & Rock, C. O. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 612–636 (ASM, Washington DC, 1996).
5. Bergler, H., Hoegenauer, G. & Turnowsky, F. *J. Gen. Microbiol.* **138**, 2093–2100 (1992).
6. Provence, D. L. & Curtiss, R. I. in *Methods for General and Molecular Bacteriology* (eds Gerhardt, P., Murray, R. G. E., Wood, W. A. & Krieg, N. R.) 317–347 (Am. Soc. Microbiol., Washington DC, 1994).
7. Singer, M. *et al. Microbiol. Rev.* **53**, 1–24 (1989).
8. Bergler, H. *et al. J. Biol. Chem.* **269**, 5493–5496 (1994).
9. Kater, M. M., Koningstein, G. M., Nijkamp, H. J. J. & Stuitje, A. R. *Plant Mol. Biol.* **25**, 771–790 (1994).
10. Rozwarski, D. A., Grant, G. A., Barton, D. H. R., Jacobs, W. R. Jr & Sacchettini, J. C. *Science* **279**, 98–102 (1998).
11. Bergler, H., Fuchsichler, S., Hoegenauer, G. & Turnowsky, F. *Eur. J. Biochem.* **242**, 689–694 (1996).



The results of this analysis indicate that the fossil taxon is nested between the genus *Petrosavia* and the family Triuridaceae, a completely achlorophyllous saprophytic clade. The clade formed by *Petrosavia*+fossil taxon+Triuridaceae is present in the all of the most parsimonious trees as well as in the consensus tree, and supports the possibility that the extinct plants were also saprophytic and achlorophyllous. We cannot be certain that the Turonian representatives of this clade were achlorophyllous, but an earlier origin of the clade, implied by the already modern aspect of the flowers and the diversity of species of Triuridaceae represented in these sediments, is consistent with this possibility. If so, this represents the earliest known fossil evidence of the saprophytic habit in angiosperms. Today, modern Triuridaceae are typically found in tropical forests, as saprophytes growing in leaf litter, or sometimes in more specialized habitats such as termite mounds.

The new fossils not only provide the oldest record of a monocotyledonous flower; they are also the most unequivocal and phylogenetically best understood of the putative early monocotyledonous fossils. The highly specialized nature of the presumed modern relatives of these monocotyledonous flowers presents a dilemma in our understanding of early monocotyledon and angiosperm diversification. Either the monocotyledonous clade is much older than its fossil representation indicates, or our interpretation of what constitutes a 'primitive' monocotyledon must be revised.

**M. A. Gandolfo, K. C. Nixon, W. L. Crepet**  
L. H. Bailey Hortorium, Cornell University,  
Ithaca, New York 14853, USA  
e-mail: mag4@cornell.edu

**D. W. Stevenson**  
New York Botanical Garden, Bronx,  
New York 10458, USA

**E. M. Friis**  
Department of Palaeobotany,  
Swedish Museum of Natural History,  
S-104 05 Stockholm, Sweden

1. Cronquist, A. *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York, 1981).
2. Daghighian, C. P. *Bot. Rev.* **47**, 517–555 (1981).
3. Herendeen, P. S. & Crane, P. R. in *Monocotyledons: Systematics and Evolution* (eds Rudall, P. J., Cribb, D. F., Cutler, D. F. & Humphries, C. J.) 1–21 (R. Bot. Gard., Kew, 1995).
4. Friis, E. M. *Ann. Missouri Bot. Gard.* **71**, 403–418 (1984).
5. Friis, E. M., Pedersen, K. R. & Crane, P. R. *Biol. Skr. Dan. Vid. Selsk. Skrift. Kgl. Vidensk.* **41**, 1–45 (1992).
6. Crane, P. R. & Herendeen, P. S. *Rev. Palaeobot. Palynol.* **90**, 319–337 (1996).
7. Crepet, W. L. *Rev. Palaeobot. Palynol.* **90**, 339–359 (1996).
8. Crepet, W. L. & Nixon, K. C. in *The Anther: Form, Function and Phylogeny* (eds D'Arcy, W. G. & Keating, R. C.) 25–57 (Cambridge Univ. Press, 1996).
9. Nixon, K. C. & Crepet, W. L. *Am. J. Bot.* **80**, 616–623 (1993).
10. Herendeen, P. S., Crepet, W. L. & Nixon, K. C. *Plant Syst. Evol.* **189**, 29–40 (1994).
11. Gandolfo, M. A., Nixon, K. C. & Crepet, W. L. *Am. J. Bot.* (in the press).
12. Davis, J. I. *Syst. Bot.* **20**, 503–527 (1995).
13. Stevenson, D. W. & Loconte, H. in *Monocotyledons: Systematics and Evolution* (eds Rudall, P. J., Cribb, D. F., Cutler, D. F. &

- Humphries, C. J.) 543–578 (R. Bot. Gard., Kew, 1995).
14. Brenner, G. J. *Bull. State Maryland, Board Nat. Resour., Dept. Geology, Mines, Water Resour.* **27**, 1–215 (1963).
15. Gandolfo, M. A. et al. *Am. J. Bot.* **84**, 483–493 (1997).
16. Gandolfo, M. A., Nixon, K. C. & Crepet, W. L. *Fifth Conf. Int. Org. Palaeobot. Abstr.* **42** (1996).

## Natural selection on human twinning

The frequency of twin deliveries varies among human populations<sup>1</sup>. The highest twinning rates for caucasian populations have been recorded on the archipelago of Åland and Åboland, in south-west Finland<sup>2,3</sup>, whereas multiple deliveries in adjacent mainland areas are historically rarer<sup>3</sup>. Using data from the pre-industrial era (1752–1850), we compare the lifetime reproductive success of mothers who produced twins with that of mothers of singletons in these archipelago and mainland sites. When we restrict our analysis to mothers with a genetic tendency to produce twins, we find that lifetime reproductive success is maximized by having twins on the archipelago, but by having singleton offspring on the mainland. This result is consistent with the difference in twinning rate being maintained by natural selection.

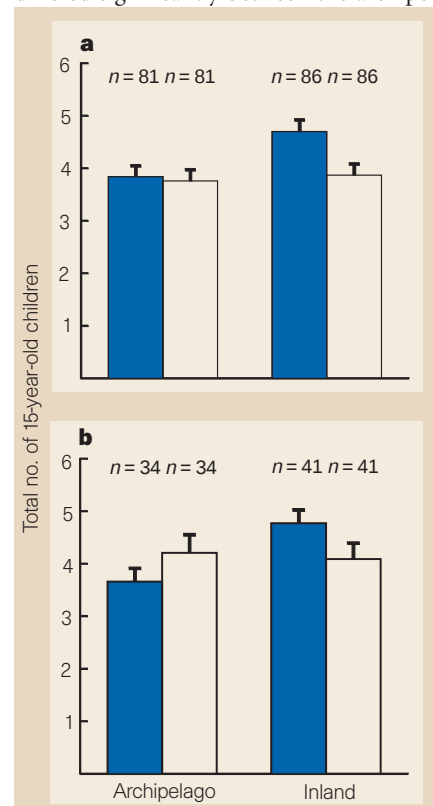
Our data are from Finnish church records of the pre-industrial period. During this era, twinning frequency was 21.3‰ on the archipelago and 14.9‰ on the mainland (comparison of twinning rates between 17 archipelago and 16 mainland parishes:  $F_{1,25} = 35.32$ ,  $P < 0.0001$ ). We collected data for eight random parishes on the mainland (Artjärvi, Ikaalinen, Haukipudas, Loimaa, Nurmijärvi, Paltamo, Sipoo and Tohmajärvi) and two on the archipelago (Rymättylä and Kustavi), and data for a further eight parishes on the mainland and fifteen on the archipelago (some small parishes were combined in the analysis) were collected by Mattila<sup>4</sup> and Eriksson<sup>3</sup>, respectively.

We analysed the reproductive output of 81 mothers of twins from the archipelago (the parishes of Hiittinen, 60° N 22° 30' E, Kustavi, 60° 30' N 21° 30' E, and Rymättylä, 60° 15' N 22° E, were used as area replicates within the archipelago) and 86 mothers of twins from the mainland (Ikaalinen, 61° 45' N 23° E, and Pulkila, 64° 15' N 26° E). We paired each mother of twins with a control mother the same age ( $\pm 3$  years) who had delivered only singleton children<sup>5</sup>. For each mother, we calculated realized lifetime reproductive output as the total number of children, including possible illegitimate children and those from earlier marriages, who had reached an age of 15 years.

Mothers of twins and singletons on the archipelago made the same contribution to the subsequent population, but on the

mainland, the reproductive output of mothers of twins was significantly lower than that of control mothers (Fig. 1a). As the figures refer to lifetime reproductive success, they include the effects of higher childbirth mortality ( $G^2 = 9.93$ , d.f. = 1,  $P = 0.0002$ ) among mothers of twins.

Because we can safely assume that only dizygotic, but not monozygotic, twinning has a genetic component<sup>6</sup>, only the analysis of dizygotic twinning is relevant to evolution of the number of children produced at one time that is driven by natural selection. Given the demographic nature of our data, we can presume that the group of mothers who delivered twins of different sexes have a higher genetic tendency towards twinning than others: all twins of opposite sexes are dizygotic, whereas twins of the same sex may be either mono- or dizygotic. Our results from a smaller data set, using only mothers with twins of different sexes, confirm the evolutionary inference: the lifetime reproductive success of mothers of twins relative to that of singleton mothers differed significantly between the archipel-



**Figure 1** Total numbers of children (+ s.e.) raised successfully to adulthood by mothers of singletons (dark bars) and mothers of twins (pale bars) in archipelago and mainland areas of pre-industrial Finland. **a**, All mothers of twins included (ANOVA: difference between areas,  $F_{1,3} = 2.46$ ,  $P = 0.21$ ; difference between mothers of twins and singletons,  $F_{1,165} = 6.11$ ,  $P = 0.015$ ; interaction between area and mother type,  $F_{1,165} = 4.47$ ,  $P = 0.036$ ). **b**, Mothers with only different-sex twins included (area,  $F_{1,3} = 1.73$ ,  $P = 0.28$ ; mother type,  $F_{1,73} = 0.08$ ,  $P = 0.77$ ; area  $\times$  mother type,  $F_{1,73} = 6.44$ ,  $P = 0.013$ ).

ago and mainland areas (Fig. 1b).

Our study provides evidence for the hypothesis that differences in twinning frequencies in historically relatively isolated human populations may be maintained by natural selection, as the differences in the profitability of twinning between the areas are consistent with the predictions of life-history models. Such models suggest that predictable resource levels favour the evolution of increased reproductive output<sup>8,9</sup>. In the archipelago, the amount of food available has traditionally been relatively high and constant, with total crop failures being rare and with survival ensured by fishing. In poor mainland areas, on the other hand, crop failures and subsequent famines have been common throughout the centuries<sup>10</sup>.

**Virpi Lummaa, Erkki Haukioja,  
Risto Lemmetyinen, Mirja Pikkola**

Section of Ecology, Department of Biology,  
University of Turku, FIN-20014 Turku, Finland  
e-mail: virpi.lummaa@utu.fi

1. Vogel, F. & Motulsky, A. G. *Human Genetics. Problems and Approaches* (Springer, Berlin, 1986).
2. Eriksson, A. W. et al. *Acta Genet. Med. Gemellol.* **37**, 277–297 (1988).
3. Eriksson, A. W. *Human Twinning In and Around the Åland Islands* (Societas Scientiarum Fennica, Helsinki, 1973).
4. Mattila, R. M. *Variation in Conceptions, Twin Frequency and Sex Ratio in Finland 1751–1969* (Kangasalan Kirjapaino, Kangasala, 1980).
5. Haukioja, E., Lemmetyinen, R. & Pikkola, M. *Am. Nat.* **133**, 572–577 (1989).
6. Parisi, P., Gatti, M., Prinzi, G. & Caperna, G. *Nature* **304**, 626–628 (1983).
7. Nevanlinna, H. R. *Hereditas* **71**, 195–236 (1972).
8. Gillespie, J. H. *Am. Nat.* **111**, 1010–1014 (1977).
9. Boyce, M. S. & Perrins, C. M. *Ecology* **68**, 142–153 (1987).
10. Soininen, A. M. *Old Traditional Agriculture in Finland in the 18th and 19th Centuries* (Forssan Kirjapaino Oy, Forssa, 1974).

## A Roman “implant” reconsidered

A supposed “wrought iron” dental implant<sup>1</sup> was recently reported from a second century CE Gallo-Roman necropolis in Chantambre (Essonne, France), but in my view the data need to be re-evaluated in the light of what is known regarding ancient and modern dentistry<sup>2–4</sup>. The item is described as “severely corroded”, for example, but an X-ray reveals a perfectly formed tooth with a smooth, intact surface free from the pitting expected on a small iron object interred for nearly 2,000 years under less than ideal conditions. The archaeological context and data on finds of iron in this and other tombs are not provided.

The production of a small, detailed replica of a human tooth in iron would test the skills of modern crafters. Less likely is that it would be accepted by a human body under questionably sterile conditions. This tooth appears to be a natural canine stained with oxides from proximity with an iron-rich

object. This explains the detailed shape, its appearance in the X-ray, and the analytical results reported<sup>1</sup>. The principal types of known ancient dental appliances fall into two categories: decorative Etruscan examples of the seventh to first centuries BC<sup>2,5,6</sup> and functional Near Eastern wire examples, developed in about 400 BC to stabilize loose teeth until they could regain natural anchorage<sup>7</sup>. Both types are known from the ancient literature and from unequivocal archaeological examples. Dental implants are unknown in the ancient medical texts or literature, and no archaeological examples have been verified.

Modern dental techniques developed late in the nineteenth century and are still evolving. The development of sophisticated implant materials that are accepted by the body is a very recent achievement<sup>8,9</sup> related to parallel research done in bone joint replacement. Dental implantology is still emerging from experimental stages<sup>10</sup>, and requires sophisticated high-technology alloys and bonds of complex composition. With space-age technology and the most modern aseptic conditions, a five-year success rate of around 85% has now been achieved. Dental loss is commonly thought of as a normal factor of ageing, with replacements being limited to the well-to-do among the most industrialized countries. The likelihood that the ancient Romans would have been interested in attempting to fashion dental implants to replace lost teeth is remote.

I therefore suggest that the Chantambre specimen is a natural tooth stained with iron oxides, and not an iron implant. We have good reason to marvel at the massive construction projects of the Romans, and at their delicate carvings on impressively hard gemstones. The ability of ancient “surgeons” throughout the world to cut pieces from human skulls and to have many of their patients survive is equally amazing. But whether they were interested in or capable of creating true dental implants in my view requires more evidence.

**Marshall Joseph Becker**

West Chester University, West Chester,  
Philadelphia 19383, USA  
e-mail: mbecker@wcupa.edu

1. Crubézy, E., Pascal Murail, L. G. & Bernadou, J.-P. *Nature* **391**, 29 (1998).
2. Becker, M. J. in *Proc. 1st Int. Conf. Anthropology and History of Health and Disease* Vol. 2 (ed. Guerri, A.) 8–19 (Erga, Genova, 1998).
3. Albrektsson, T., Zarb, G. A., Worthington, P. & Eriksson, A. R. *Int. J. Oral Maxillofacial Implants* **1**, 11–25 (1986).
4. Becker, M. J. *Dental Anthropol. Newsl.* **9**, 5–10 (1994).
5. Becker, M. J. *J. Paleopathol.* **6**, 69–92 (1994).
6. Becker, M. J. in *Actas del I Congreso Internacional de Estudios sobre Momias, 1992* (ed. Aufderheide, A. C.) Vol. 2, 651–658 (Museo Arqueológico y Etnológico de Tenerife, Santa Cruz de Tenerife, 1995).
7. Becker, M. J. *Berytus* **42**, 71–102 (1995/6).
8. Schnitman, P. A. & Schulman, L. B. (eds) *Dental Implants. Benefits and Risk*. (U.S. Dept of Health and Human Services, Washington, 1979).
9. Braonemark, P.-I. et al. *Scand. J. Plastic Reconstruct. Surg.*

Suppl. 16 (1977).

10. Henry, P. J. et al. *Int. J. Oral Maxillofacial Implants* **11**, 450–455 (1996).

**Crubézy et al. reply** — We disagree with Becker’s view that the dental implant described in our earlier Scientific Correspondence<sup>1</sup> is a natural canine stained with iron oxides. The dental implant was located in a position normally taken by the upper second right premolar, a position in which a normal canine would not be found. Furthermore, the only goods associated with this burial were pottery, not iron or any metal objects<sup>2</sup>. Even if there had been iron oxide contamination, it is unlikely that it would have affected only one tooth. Figure 1a in our earlier Scientific Correspondence shows that the piece of metal is corroded on its periphery; the “smooth, intact surface” observed on the X-ray is a common artefact of the technique. Finally, we have already noted that the implant was broken and that metallurgical analysis unambiguously identifies it as metal and not as a biological tissue.

The fabrication of a “detailed replica” of a human tooth is not as dubious as Becker maintains. Chemical analysis indicates that the metal was given its shape through hot-hammering and folding, a basic technique of ancient blacksmiths, including those of Gallo-Roman times. Concerning the successful retention of the implant, it is possible that the iron could have facilitated the osseointegration<sup>3</sup>; the absence of aseptic conditions does not systematically imply the rejection of the implant. The success of this procedure in an ancient population is no more amazing than the 70 per cent survival rate among patients who underwent trepanation<sup>4</sup> or the successful performance of cataract surgery<sup>5</sup>.

Thus our anatomical, morphological, metallurgical and microscopic analyses of this specimen document, without question, the successful implantation of this dental prosthesis.

**Eric Crubézy**

Fédération d’Anthropologie,  
Université Paul Sabatier,  
31 000 Toulouse, France

**Pascal Murail, Louis Girard**

Laboratoire d’Anthropologie, Avenue des facultés  
33 405 Talence cedex, France

**Jean-Pierre Bernadou**

Ecole Nationale Supérieure de l’Aéronautique et de l’Espace, 31 055 Toulouse, France

1. Crubézy, E., Murail, P., Girard, L. & Bernadou, J. P. *Nature* **391**, 29 (1998).
2. Murail, P. *Biologie et Pratiques Funéraires des Populations d’Epoque Historique : une Démarche Méthodologique Appliquée à la Nécropole Gallo-Romaine de Chantambre (France)*. (Thesis, Bordeaux Univ., 1996).
3. Claes, L., Hutzschenreuter, P. & Pohler, O. *Arch. Orthop. Unfallchir* **85**, 155–163 (1976).
4. Stewart, T. D. *Annual Report of the Smithsonian Institution* 469–491 (1958).
5. Feugère, M., Kunz, E. & Weisser, U. *Dossiers d’Histoire et d’Archéologie* **123**, 66–71 (1988).

# Focusing on what counts

## The Man Who Loved Only Numbers: The Story of Paul Erdős and the Search for Mathematical Truth

by Paul Hoffman

Fourth Estate: 1998. Pp. 302. £12.99, \$22.95

Alexander Masters

"A mathematician," remarked Paul Erdős, "is a machine for turning coffee into theorems." Erdős's particular passion was for prime numbers. He would march up and down the street flapping his arms in excitement as he thought about them. He was in awe of their fundamental importance; he marvelled at their superficial simplicity. When it comes to primes, said Erdős, "babies can ask questions that grown up men can't answer". (Hence, the reason child prodigies so often excel at number theory.) For example, Goldbach's Theorem. In 1742, Goldbach noted that every even number he tested was the sum of two primes:

$4=2+2$ ;  $6=3+3$ ;  $8=5+3$ ;  $10=5+5$ ; etc.

Nobody yet knows, 250 years later, whether or not the theorem is universally true.

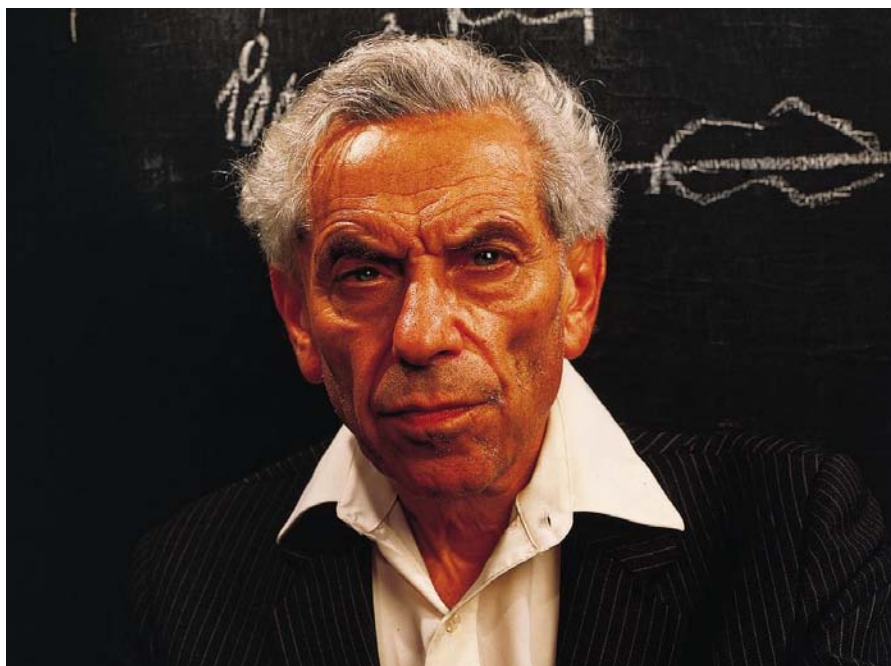
Erdős's first triumph, published as a first-year undergraduate, was a new proof that a prime number can always be found between any integer and its double. The great Chebyshev had published the first proof in 1850, but it was a clumsy thing. A good proof, according to G. H. Hardy, has "a very high degree of *unexpectedness*, combined with *inevitability* and *economy*. The argument takes so odd and surprising a form; the weapons used seem so childishly simple when compared with the far-reaching consequences; but there is no escape from the conclusions". Chebyshev "had used a steam shovel to transplant a rosebush", remarks Paul Hoffman in one of the many felicitous metaphors in *The Man Who Loved Only Numbers*. "Erdős managed it with a spoon." His coup was celebrated in rhyme throughout the mathematical world:

*Chebyshev said, and I say it again,*

*There is always a prime between  $n$  and  $2n$ .*

Erdős (pronounced 'air-dish') was a peculiar man. He was small and frenetic, and wore only silk, including silk underwear and silk socks, because of unspecified skin complaints. His disdain for sex and eroticism was famous. He put it down to two factors: first, he had always wished to be different, and when the other boys at school were making fools of themselves with women, he determined to have none of it. Second, he had a physical complaint involving the organ in question; it was curable, but he could never afford enough time off from his mathematics to get it fixed.

He lived out of two half-empty suitcases, travelling from university to university, staying in this hotel or with that friend, and giv-



Soft as silk: Erdős was an impossible guest, but gave away his money and helped other researchers.

ing away all his prize money and income. His work popped up in many languages, in all sorts of obscure journals:

*A conjecture both deep and profound*

*Is whether the circle is round.*

*In a paper of Erdős*

*Written in Kurdish*

*A counterexample is found.*

As a house guest Erdős was a thorough nuisance. He took 10–20mg Bensedrine or Ritalin every day, backed up by caffeine tablets and espresso, and worked for 19 hours. He would keep his host awake until 2am, and then get up again at 4.30am and crash the pots about the kitchen to signify that he'd had enough sleep and it was time to start talking mathematics again. "Dear So-and-so," a typical letter from him would begin. "I am in Australia. Tomorrow I leave for Hungary. Let  $k$  be the largest integer..."

Hoffman is very good at discussing the sort of problems that intrigued Erdős, which makes *The Man Who Loved Only Numbers* an excellent introductory book for anyone interested in number theory. As a subject, however, Erdős is difficult. He led a determinedly undramatic life and Hoffman is forced to fill up many pages with unrelated detail and stories about other mathematicians. This upsets the flow of the biography, but frequently adds to the book's interest. There is, for example, a long aside about Fermat, and the Last Theorem, and about Andrew Wiles and his proof. Erdős, whose interests were comparatively narrow, did not understand the proof, and disapproved of Wiles's furtive attempts to make sure that no

one else shared in the discovery. There are other intriguing sections about Cantor and Gödel (both considerably more eccentric than Erdős), and a whole, slightly woolly, chapter about Erdős's close friend Ron Graham. Mr Graham was Hoffman's most helpful source and gets four photos to himself in the plates.

Although Erdős confined himself largely to number theory, he often dabbled in other fields. Hoffman quotes a mathematician who, written on his blackboard, had a thorny problem in functional analysis, a field Erdős knew nothing about. A 30-page solution had just been discovered, and the authors were feeling proud about it. Erdős went up to the board and said: "What is that? Is it a problem?" "Yes," said the mathematician. "What do the symbols mean?" The mathematician explained them to him. Erdős then immediately wrote down a two line solution. "If that's not magic," said the man afterwards, "then what is?"

Erdős was also remarkably good at inventing problems on the spur of the moment. They were always designed to be just beyond a person's current reach, and often had a reward attached to them. "How many people must have got started on research by solving a \$5, or maybe even a \$1, Erdős problem?" recalled one friend. He made it his business to push forward child prodigies, give confidence to new research students and return confidence to mathematicians who had started to doubt their abilities. If any mathematician gave up the subject, Erdős said that they had "died".



In old age, when he was threatened with blindness in one eye, he had to be forced to have a corneal transplant to save his sight.

"Doctor," said Erdős, "will I be able to read?" "Yes," said the doctor. "That's the whole point of the surgery." In the operating room, Erdős immediately demanded to know why the doctor had turned the lights down. "So we can do the surgery." "But you said I'd be able to read." The doctor then had to call the mathematics department at the university and have a professor sent over so that Erdős could talk about mathematics during the operation.

*The Man Who Loved Only Numbers* is a fascinating, affectionate biography, but the title is untrue: Erdős loved many things besides numbers, including children (whom he called epsilons, after the mathematical symbol for a very small quantity) and especially his mother, who accompanied him on many of his early travels. He died in 1996. His ashes are buried alongside her, in Hungary. □

Alexander Masters is at 10 Hertford Street, Cambridge CB4 3AG, UK.

## The making of a bomb scientist

### Peace and War: Reminiscences of a Life on the Frontiers of Science

by Robert Serber with Robert P. Crease  
Columbia University Press: 1998. Pp. 241.  
\$29.95, £24

Richard Rhodes

A shy and unassuming theoretical physicist, Robert Serber was responsible for the design of the Hiroshima bomb. Born in Philadelphia in 1909, he was a protégé of Robert Oppenheimer, whom he met at the University of Michigan's famous physics summer school in Ann Arbor in 1934.

Serber had taken his PhD the previous spring (having delayed as long as possible to



"Our manufactured hell": Serber, centre, assessing bomb damage among the ruins of Nagasaki.

avoid leaving graduate school for Depression unemployment) and was preparing to study at Princeton with Eugene Wigner on a post-doctoral fellowship. After the meeting in Ann Arbor, he decided to follow Oppenheimer to the University of California at Berkeley instead. They published eight papers together and were close friends by the time Oppenheimer accepted an appointment with the Manhattan Project as coordinator of rapid rupture and, later, director of the Los Alamos laboratory.

Serber had moved to the University of Illinois at Urbana when his Berkeley colleague recruited him for bomb work. At the secret conference Oppenheimer organized in Berkeley in summer 1942 — where theorists Hans Bethe, Edward Teller, Richard Tolman, Emil Konopinski, Felix Bloch and others reviewed the bomb's feasibility — Serber continued the work on a fission weapon while Teller diverted the rest of the group with thermonuclear visions. Even implosion had its origins in discussions between Serber and Tolman at Berkeley that summer.

Inevitably, then, it was Serber who delivered the opening technical lectures at Los Alamos in April 1943 which, written up by Edward Condon, became the famous *Los Alamos Primer*.

"A couple of minutes into the first lecture," Serber recalls in *Peace and War*, "Oppie sent John Manley up to tell me not to use the word 'bomb' but to use something neutral like 'gadget'. I did, the word stuck, and after that the bomb was always called 'the gadget' at Los Alamos."

Serber also named the gun design Thin Man, after a recent movie based on the Dashiell Hammett detective mystery (it shrank to Little Boy when the barrel was shortened), and the implosion design Fat Man, after the portly character actor Sidney Greenstreet.

After the first bomb was tested at the Trinity site in New Mexico in July 1945, Serber went out to Tinian, in the Mariana islands, to assist with weapons assembly. He was scheduled to run a high-speed Fastax camera to photograph the Nagasaki bombing from the air, but the B-29 pilot ejected him at the head of the runway because the crew was short of a parachute, and the footage was lost. In September, he travelled with William Penney and other Los Alamos colleagues to Nagasaki and then Hiroshima to assess bomb damage.

Recent Japanese estimates of the atomic bombing death tolls give 140,000 for Nagasaki. But on 13 September 1945, Serber interviewed a Japanese naval intelligence officer in Nagasaki, whose count is now reported in *Peace and War*: 19,743 dead, 1,927 missing and 40,993 wounded. Serber had encountered emaciated Allied prisoners of war at Nagasaki and burned Japanese bomb victims; he was neither exultant nor apologetic about the bombings. In a contemporary letter to his wife, Charlotte, he called the damage at Nagasaki "our manufactured hell".



Ground zero: office safes were among the few items left standing after the bomb fell on Nagasaki.



After the war, Serber worked with Ernest Lawrence at the Radiation Laboratory at Berkeley. Lawrence sent him to Washington in October 1949 to represent the laboratory in the secret debate proceeding in the United States government about whether to hasten development of a hydrogen bomb. He was surprised to find Oppenheimer doubtful and more senior scientific advisers such as James Conant and I. I. Rabi adamantly opposed. The debate, and its odious fallout in the 1954 Oppenheimer security hearing, deeply divided the physics community.

One of Serber's saddest stories concerns this time. Oppenheimer's attorney asked Serber not to communicate with his old friend during the security investigation because previous (unfounded) accusations of disloyalty against Serber and his wife might re-emerge to cause Oppenheimer harm. Only years afterwards did Serber learn that Oppenheimer had not authorized the attorney's call and had been hurt by Serber's silence. No one, not even Teller, benefited from the destruction of Oppenheimer.

Serber's contribution to physics included his participation with Oppenheimer in the theoretical work that led to the discovery of mesons. In 1951 he moved to Columbia University at Rabi's urging ("You have to choose between Ernest and Oppie," he says Rabi told him). There he deduced the existence of the elementary particles inside baryons and mesons that came to be called quarks. "I took [Murray Gell-Mann] to lunch at Columbia's Faculty Club," he writes, "and explained this idea to him." Serber didn't quite believe it, however — he remembers thinking the fractional charges were "appalling" — but saw the light a day or two later when he calculated the magnetic moments. "At that point," he writes, "I should have published; but I never got around to it."

A friend thought he procrastinated because the idea seemed obvious to him (a common hurdle in the development of original ideas). Needless to say, Serber regretted not having published his insight; he receives an acknowledgement in Gell-Mann's paper.

Serber's well-told reminiscences originated as the 1994 George B. Pegram Lectures at Columbia. Robert P. Crease, a philosopher and historian of science, worked with Serber before the physicist's death in 1997 to expand them skilfully and unobtrusively into this engaging memoir. Serber experienced personally and contributed vitally to the historic confrontation between science and the nation-state that may be the century's most significant legacy: the application of nuclear energy to the development of weapons too destructive to risk using, the *reductio ad absurdum* of national power. I learned much from this book. □

Richard Rhodes, an independent journalist and historian, is at 609 Summer Hill Road, Madison, Connecticut 06443, USA.

## NO sex please...

**NO**

by Carl Djerassi

University of Georgia Press: 1998. Pp. 276.

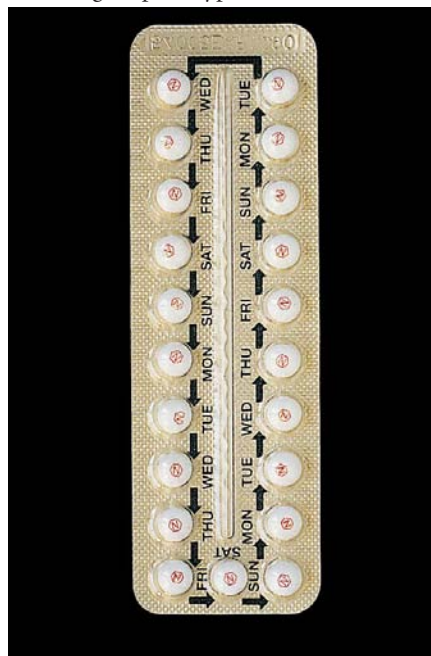
\$24.95

**Frances M. Brodsky**

*NO* is the fourth novel in Carl Djerassi's tetralogy of "science-in-fiction". Extremely topical, given the current Viagra-mania, the story of *NO* evolves around the scientific and biotechnological applications of the multifunctional gas, nitric oxide, infamous for its regulatory role in penile erection, as well as causing the yawning reflex in rats.

In the proselytizing preface, Djerassi describes his socio-scientific goals in writing *NO* and reveals that the book draws liberally on his own experiences in the academic and biotechnological world of sex science, as professor of chemistry at Stanford and the organic chemist behind the development of the contraceptive pill.

In his laudable campaign to expand the genre of "science-in-fiction", a topic on which he has written in *Nature* and teaches at Stanford (see *Nature* 392, 244 & 393, 511–513; 1998), Djerassi joins a respectable tradition, well represented in Walter Gratzer's inspiring anthology *A Literary Companion to Science* (Longman, 1989; Norton, 1990). Gratzer's book is required reading for scientists curious about their image in the literary world. It comprises portrayals of science by non-scientific writers, such as John Updike's wry description of a Harvard laboratory in the midst of his otherwise soft-pornographic novel *Couples*, as well as excerpts from scientist authors of fiction, including the prototype, C. P. Snow.

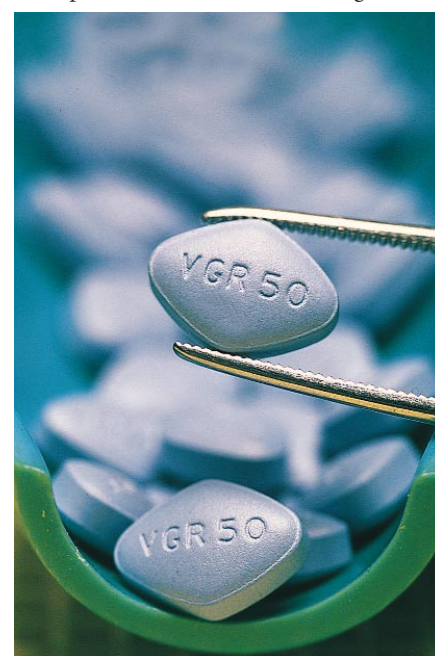


Snow's fictional attempt to bridge the "two cultures", in his series *Strangers and Brothers* (1947–70), is still relevant to modern issues of science, ethics and society, and is a model of how an author can portray scientists as human, while effectively explaining their enterprise. These are the goals that Djerassi, and others aspiring to the genre, have set themselves.

But the results can only be as good as the fiction itself. As with many novelists who are scientists, Djerassi occasionally lapses into scientific mode, stating the point he would like to make, rather than revealing it through the actions and relationships of his characters. *NO* traces the scientific career of Dr Renu Krishnan, a woman of Indian background, whose research leads her to synthesize substances that release nitric oxide, and who gets drawn into the commercial development of these compounds for the treatment of so-called erectile dysfunction. In general, Djerassi plausibly renders the point of view of a woman from an ethnic minority, although he does trip on a few details. The conflict that Djerassi paints between Krishnan's Indian background and that of her Jewish Israeli lover quite effectively brings out the repressive aspects of both cultures.

Interestingly, Krishnan is most out of character when giving a scientific presentation. She becomes unconvincingly savvy and calculating, suspiciously resembling the author himself. Djerassi, by the way, makes several cameo appearances in the novel.

The initial interactions of Krishnan with the male scientific world, discussing treatment of erectile dysfunction, are exploited for their humour, as well as for inherent social commentary. Djerassi excels in locker-room jokes, which are ostensibly blessed with political correctness, coming as fre-



**Thrills and pills:** Carl Djerassi's invention, the oral contraceptive, and the impotence drug Viagra.

quently from Krishnan as from her nemeses.

But after the first 90 pages, apart from a mildly interesting sub-plot involving sperm stealing, the novel takes a more mundane turn and becomes a description of the biotechnology business plan to capitalize on Krishnan's discoveries. The interesting human aspects of the characters, the inner conflicts of scientists succumbing to competitive drives and the temptations of commercialization, become secondary to the not-so-suspenseful fate of their stock options.

Overall, the reader is most likely to be gripped by the well-researched biology of NO and the "jouissance" derived from reading about the science of sex, a term that, according to Djerassi, was fashionable among undergraduates at Wellesley College, Massachusetts, circa 1970. □

Frances M. Brodsky is in the Departments of Biopharmaceutical Sciences, Pharmaceutical Chemistry and Microbiology and Immunology at the University of California, 513 Parnassus Avenue, San Francisco, California 94143-0552, USA. She is the author of the scientific mystery *Principal Investigation* by B. B. Jordan (*Berkley Prime Crime*, 1997).

## From chaos to complexity

### Chaos Theory Tamed

by Garnett P. Williams

Taylor & Francis: 1997. Pp. 499. £19.95, \$34.95

### Dynamics of Complex Systems

by Yaneer Bar-Yam

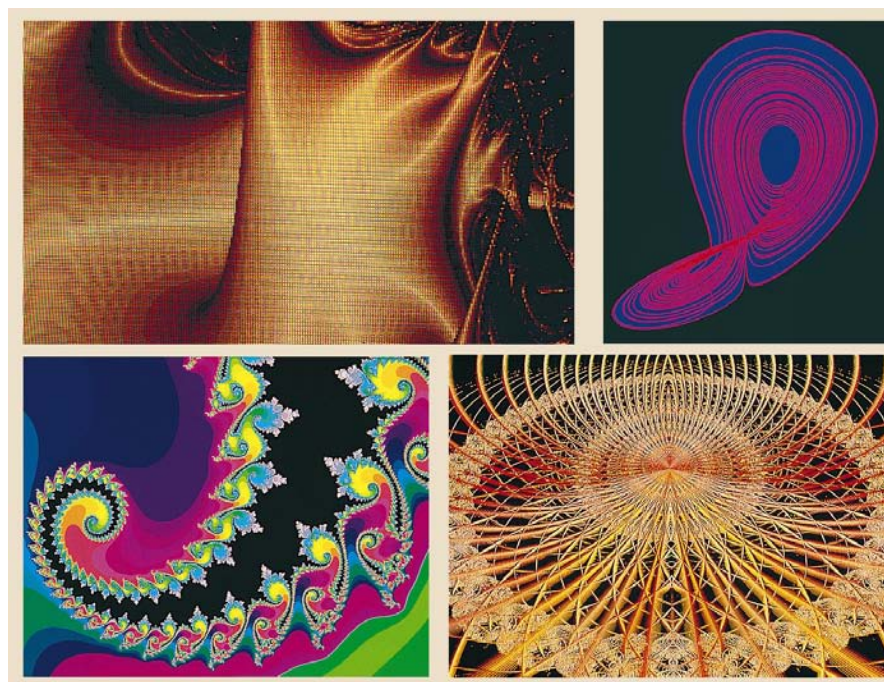
Addison-Wesley: 1997. Pp. 848. \$56

Michael F. Shlesinger

Chaos is no longer a new field. It has already been 35 years, and several generations of students, since Edward Lorenz discovered the strange attractor. Neither is chaos a fad or a dead end. It is based on the rock-solid foundation of physics, Newton's laws, and on tackling the nonlinear, non-integrable equations whose solution had to wait for an appreciation of unstable behaviour, new mathematical tools and the advent of computer visualization.

The thesis of Garnett Williams's *Chaos Theory Tamed* is that enough wisdom has accumulated to give an account of chaos theory mostly in words and pictures, without resorting to deep and sophisticated mathematics.

Williams is careful to focus on standard theoretical topics related to low-dimensional dissipative systems, and opts not to broach the rich, complex subject of Hamiltonian systems, thereby omitting topics such as the three-body problem and the strange kinetics associated with the fractal



Out of the chaos: clockwise from top left are a Lyapunov space (used to study how enzymes break down carbohydrates), the Lorenz Attractor and fractal images entitled *Overseer* and *Scorpio's Tail*.

orbits-within-orbits of the standard and Zaslavsky maps.

Williams succeeds in his goal, with a carefully written, thoughtful exposition of standard topics (such as the logistic map, strange attractors, routes to chaos and Poincaré sections) and tools of the trade (including attractor reconstruction, Kolmogorov–Sinai entropy, fractal dimensions and Lyapunov exponents). Equations are included, but they are developed using careful discussion, rather than detailed mathematics. The first 158 pages of his book cover background information, including vectors, Fourier analysis, probability theory and time series. A good deal of discussion is given to the logistic map as a simple model to demonstrate many ideas of chaos. Those unfamiliar with the basic topics of nonlinear dynamics in dissipative systems would do well to study this friendly, self-contained book.

The success of nonlinear dynamics in handling chaos in systems with few degrees of freedom has led some to believe that these methods can be extended all the way to understanding complex social systems, such as economics, war strategy, psychology and city planning. But the ultimate success of the ideas of chaos in physics has been based on experimental verification of the existence of nonlinear instabilities and behaviours in well-controlled, repeatable experiments — in other words, the scientific method. Like the proverbial river into which one cannot step twice, one cannot repeat a social experiment, because the first experiment changes the conditions under which it was executed. But much can be

learned, and a similar limitation has not deterred cosmologists.

Yaneer Bar-Yam's intriguing *Dynamics of Complex Systems* goes beyond chaos theory to the broader field of complex systems. He does not define complexity, but considers mainly systems with a large number of interacting parts, and seeks to discover pervading themes, such as memory, adaptation, evolution and self-organization, and then to model these phenomena.

The book begins with a 294-page introduction — a veritable book within a book — covering basic topics such as iterative maps, Monte Carlo techniques, random walks, phase transitions, activated processes and fractals. These topics form an extensive toolkit, providing the reader with the means to characterize, model and simulate aspects of complex systems.

In the body of the book, Bar-Yam begins with neural networks, then moves up the scale of complexity to protein folding, evolution, developmental biology and, finally, human civilization. The book does not try to have the last word on these vast fields, but introduces the reader to aspects that can be modelled and explored. Throughout, questions and their answers are folded into the text, and the many mathematical techniques and arguments are clearly presented. This book is an excellent place to start exploring the concepts and techniques of complex systems and provides an effective springboard to further studies. □

Michael F. Shlesinger is in the Office of Naval Research, Physical Sciences Division 331, 800 North Quincy Street, Arlington, Virginia 22217-5660, USA.



# Design and self-assembly of two-dimensional DNA crystals

Erik Winfree\*, Furong Liu†, Lisa A. Wenzler† & Nadrian C. Seeman†

\* Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125, USA

† Department of Chemistry, New York University, New York, New York 10003, USA

**Molecular self-assembly presents a 'bottom-up' approach to the fabrication of objects specified with nanometre precision. DNA molecular structures and intermolecular interactions are particularly amenable to the design and synthesis of complex molecular objects. We report the design and observation of two-dimensional crystalline forms of DNA that self-assemble from synthetic DNA double-crossover molecules. Intermolecular interactions between the structural units are programmed by the design of 'sticky ends' that associate according to Watson–Crick complementarity, enabling us to create specific periodic patterns on the nanometre scale. The patterned crystals have been visualized by atomic force microscopy.**

Control of the detailed structure of matter on the finest possible scale is a major goal of chemistry, materials science and nanotechnology. This goal may be approached in two steps: first, the construction of individual molecules through synthetic chemistry; and second, the arrangement of molecular building blocks into larger structures. The simplest arrangements of molecular units, in two or three dimensions, are periodic—a crystal. Design components for crystals must have definable intermolecular interactions and must be rigid enough to prevent the formation of ill-defined aggregates<sup>1</sup>. Branched DNA molecules with sticky ends are promising for macromolecular crystal design<sup>2</sup> because their intermolecular interactions can be programmed through sticky ends<sup>3</sup> that associate to produce B-form DNA<sup>4</sup>; however, studies of three- and four-arm junctions reveal that the angles flanking their branchpoints are flexible<sup>5,6</sup>.

The need for a rigid design component with predictable and controllable interactions has led to the utilization of the antiparallel DNA double-crossover motif<sup>7</sup> for this purpose. Double-crossover (DX) molecules are analogues of intermediates in meiosis<sup>8</sup> that consist of two side-by-side double-stranded helices linked at two crossover junctions. Antiparallel DX molecules have a rigidity that is lacking in conventional branched junctions, indicating that they might be suitable for use in the assembly of periodic matter<sup>9</sup>.

These findings stimulated a theoretical proposal to use two-dimensional (2-D) lattices of DX molecules<sup>10</sup> for DNA-based computation<sup>11</sup>. In the mathematical theory of tilings<sup>12</sup>, rectangular tiles with programmable interactions, known as Wang tiles, can be designed so that their assembly must mimic the operation of a chosen Turing Machine<sup>13</sup>. DX molecules acting as molecular Wang tiles could self-assemble to perform desired computations<sup>10,14,15</sup>. Consequently, the ability to create 2-D lattices of DX molecules assumes additional interest as a step towards the design of molecular algorithms.

Here we report the assembly, from DX molecules, of three 2-D lattices with two distinct topologies. The DX molecules,  $\sim 2 \times 4 \times 13$  or  $2 \times 4 \times 16$  nm in size, self-assemble in solution to form single-domain crystals as large as  $2 \times 8 \mu\text{m}$  with uniform thickness between 1 and 2 nm, as visualized by atomic force microscopy<sup>16</sup> (AFM). By incorporating a DNA hairpin into a DX molecule to serve as a topographic label, we have produced stripes above the surface at intervals of 25, 32 and 64 nm. Two-component lattices have been assembled with a stripe every other unit; by programming sticky-ended associations differently, four-component lattices have been produced with a stripe every fourth unit. This observation demonstrates that it is possible to create specific lattices

with programmable structures and features on a nanometre scale.

## Design of DNA crystal

Our approach to two-dimensional crystal design is derived from the mathematical theory of tiling<sup>10,12</sup>. The desired lattice is specified by a set of Wang tiles with coloured edges; the Wang tiles may be placed next to each other only if their edges are identically coloured where they touch (Fig. 1a). Our goal is to design synthetic molecular units corresponding to these tiles, such that they will self-assemble into a crystal that obeys the colouring conditions. As an initial demonstration of molecular Wang tiles, we have chosen the simplest non-trivial set of tiles: two tiles, A and B, which make a striped lattice (Fig. 1a, left). We also demonstrate a set of four tiles that produce a striped lattice with a greater period (Fig. 1a, right). Translated into molecular terms, we obtain DX systems that self-assemble in solution into two-dimensional crystals with a well defined subunit structure.

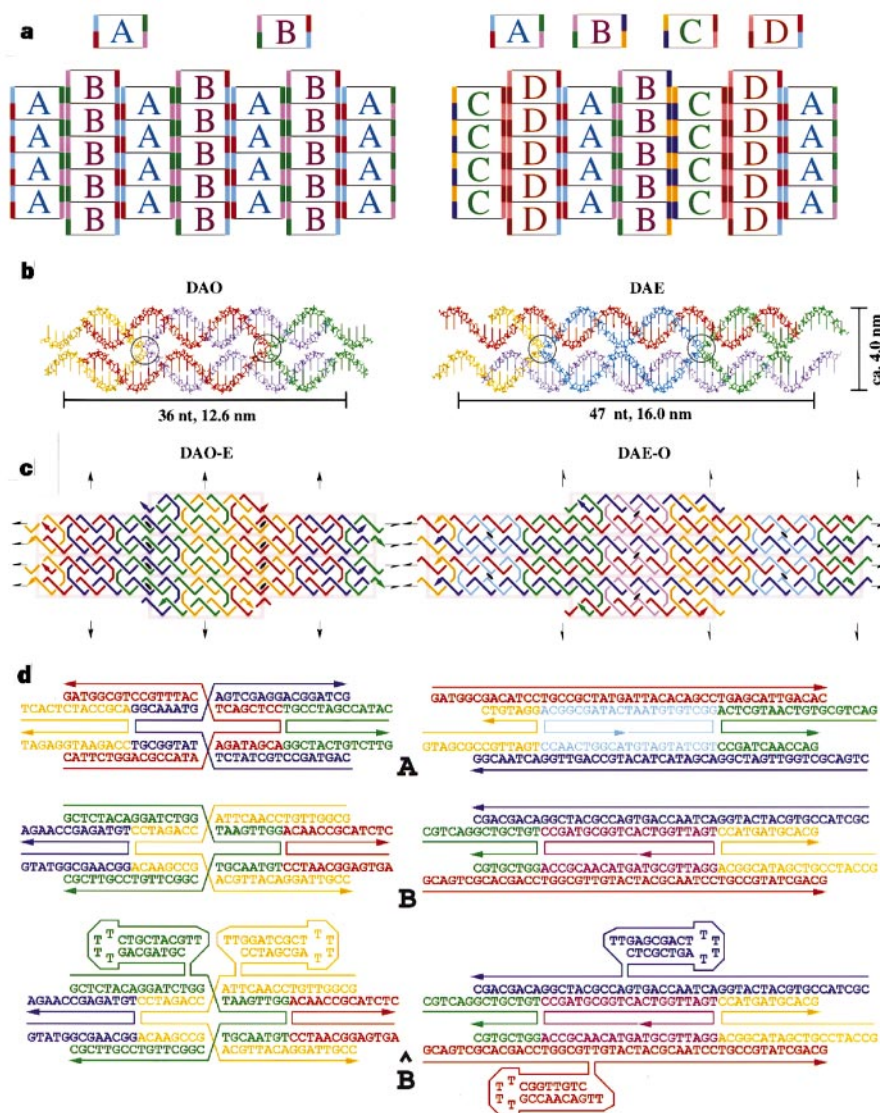
The antiparallel DX motif<sup>7</sup> consists of two juxtaposed immobile 4-arm junctions<sup>2</sup> arranged so that at each junction the non-crossover strands are antiparallel to each other. There are five distinct DX motifs, but only two are stable in small molecules<sup>7</sup>: these are called DAO (double crossover, antiparallel, odd spacing) and DAE (double crossover, antiparallel, even spacing). The design depends critically upon the twist of the B-form DNA double helix, in which a full turn takes place in  $\sim 10.5$  base pairs<sup>17,18</sup>. DAO molecules have an odd number of half-turns (e.g. 3 half-turns is  $\sim 16$  base pairs) between the crossover points, whereas DAE molecules have an even number of half-turns (for example, 4 half-turns is  $\sim 21$  base pairs). Computer models of the DX molecules used in this study are shown in Fig. 1b. The DAO molecules consist of four strands of DNA, each of which participates in both helices. The DAE molecules consist of three strands that participate in both helices (yellow, light blue, green), and two strands that do not cross over (red, dark blue). Each corner of each DX unit has a single-stranded sticky end with a unique sequence; specific association of DX units is controlled by choosing sticky ends with Watson–Crick complementarity.

To ensure that the component strands form the desired complexes, strand sequences must be designed carefully so that alternative associations and conformations are unlikely. Therefore we must solve the 'negative design problem'<sup>19–21</sup> for DNA: find sequences that maximize the free energy difference between the desired conformation and all other possible conformations. We use the heuristic principle of sequence symmetry minimization<sup>2,19</sup> to minimize the length and number of unintentional Watson–Crick

complementary subsequences. In each DX molecule's sequences, there are no 6-base subsequences complementary to other 6-base subsequences except as required by the design, and spurious 5-base complementarity is rare. Thus it is expected that during self-assembly, the DNA strands spend little time in undesired associations and form DX units with high yield.

DX units can be designed that will fit together into a two-dimensional crystalline lattice. Here we use either two or four distinct unit types (Fig. 1a) to produce striped lattices. We use two separate systems to implement the two-unit lattice, one consisting of two DAO units, the other consisting of two DAE units. The

lattices produced by these systems are called DAO-E and DAE-O, respectively, to indicate the number of half-turns between crossover points on adjacent units; their distinct topologies are shown in Fig. 1c. Covalently joining adjacent nucleotides at nicks in the lattice, by chemical or enzymatic ligation, would result in a 'woven fabric' of DNA strands. Ligation of the DAO-E design produces four distinct strand types, each of which continues infinitely in the vertical direction (here, the terms 'vertical' and 'horizontal' will always be as in Fig. 1c). The DAE-O design involves two small nicked circular strands in addition to four infinite strands, two of which extend horizontally and two of which extend vertically. The DAO-E



**Figure 1** Design of DX molecular structure and arrangement into 2-D lattices. **a**, The logical structure for 2-D lattices consisting of two units and four units. In the two-unit design, type **A** units have four coloured edge regions, each of which match exactly one coloured region of the adjacent type **B** units. Similarly, in the four-unit design, the edge colours are chosen uniquely to define the desired relations between neighbouring tiles. Note that rotations and reflections of Wang tiles are disallowed; an equivalent restriction could also be obtained by using non-rectangular tiles or more complex patterns of colours. **b**, Model structures for DAO and DAE type **A** units. Each component oligonucleotide is shown in a unique colour. The crossover points are circled. Complete base stacking at the crossover points is assumed. Computer models showing every nucleotide (nt) were generated using NAMOT2 (ref. 39). **c**, The lattice topologies produced by the DAO and DAE units. Each DX unit is highlighted by a grey rectangle. A unique colour is

chosen for each strand type which would be formed after covalent ligation of units. Arrowheads indicate the 3' ends of strands. Black ellipses indicate dyad symmetry axes perpendicular to the plane; black arrows indicate dyad axes in the plane (full arrowhead) or screw axes (half arrowhead). The symmetries of the DAO-E and DAE-O lattices are those corresponding to the layer groups<sup>40</sup>  $P2_12_2$  and  $P2_12_12_1$  respectively. The boundaries of the DAE-O units are not designed to coincide with the vertical symmetry elements. **d**, The actual sequences used in the reported experiments (see Methods for several exceptions). The schematics accurately report primary and secondary structure—oligonucleotide sequence and paired bases—but are not geometrically or topologically faithful because they do not show the double-helical twist. Both type **B** and type  $\bar{B}$  are shown, indicating where the hairpin sequences are inserted. Note that for each type—**A**, **B**, and  $\bar{B}$ —there is both a DAO unit and a DAE unit.

design has the advantage of using simple 4-strand DX units, whereas in the DAE-O design, the horizontal and vertical strands can serve as reporters for the extent of self-assembly if the molecules are ligated and analysed by gel electrophoresis. We have investigated both the DAO-E and DAE-O systems in parallel to show that the same principles of self-assembly govern both systems. Therefore, except where noted, our discussion will apply to both systems.

Control of self-assembly to yield the 2-D lattice is obtained by two design criteria. First, the sticky-end sequences for each desired contact are unique; this ensures that the orientations and adjacency relations of the DX units comply exactly with the design in Fig. 1a. Sticky ends are lengths 5 and 6 for the DAO-E and DAE-O designs respectively, so that each correct contact contributes  $\sim 8$  or  $14 \text{ kcal mol}^{-1}$  to the free energy of association at  $25^\circ\text{C}$ , according to a nearest-neighbour model<sup>22</sup>. Second, the lengths of the DX arms and sticky ends, and thus the separations between crossover points, respect as closely as possible the natural twist of the B-form DNA double helix; thus adjacent DX molecules are effectively coupled by torsional springs whose equilibrium positions have been designed to keep the adjacent DX molecules coplanar. For example, a linear rather than planar polymer could result if each unit makes two sticky-end bonds to each neighbouring unit, but this would require overtwisting, undertwisting, or bending of the double helix, and thus is discouraged by our design.

Figure 1d shows the DX units and sequences used in our experiments, except as noted in Methods. In each system, there are two fundamental DX units, called **A** and **B**, and, additionally, an alternative form **B** that contains two hairpin-terminated bulged 3-arm junctions (similar to the DX + J motif<sup>9</sup>). On the basis of studies of bulged 3-arm junctions<sup>23</sup>, we expect that in each unit, one hairpin will point up and out of the plane of the DX crystal, while the other hairpin will point down and into the plane, without significantly affecting the rigidity of the molecule<sup>9</sup>. The **B** units will replace the **B** units and serve as contrast agents for AFM imaging, because their increased height can be measured directly. Related sequences were used for the four-unit lattice **ABCD** and for studies involving gold labelling or alternative placements of the hairpins.

### Characterization by gel electrophoresis

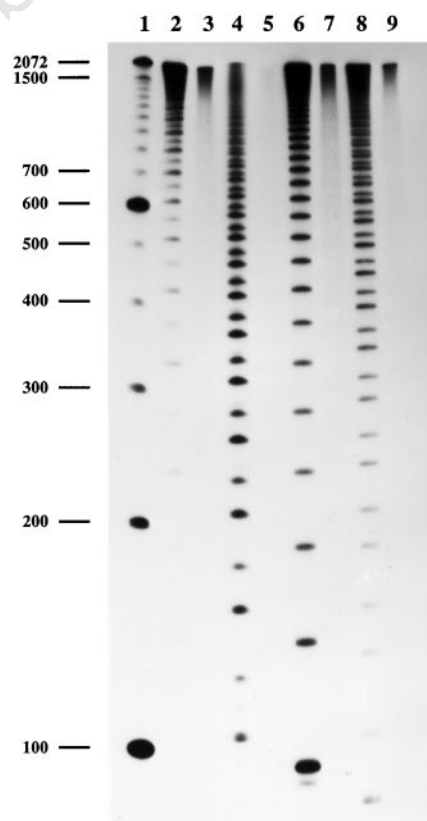
A prerequisite for lattice self-assembly is the formation of the DX units from their component strands. A thorough investigation of this issue was done for the original studies of DX<sup>7</sup>; that the new designs also behave well constitutes further validation of the antiparallel DX motif. Because the sticky ends of **A** units have affinity only for sticky ends of **B** units, and not for themselves, neither **A** nor **B** alone in solution can assemble into a lattice. Thus the formation of isolated DX units can be monitored easily by non-denaturing gel electrophoresis, as described previously<sup>7</sup>, and more than 95% of the material is seen in the expected band.

Solutions containing **A** units and **B** units can be mixed and annealed to form **AB** lattices. Enzymatic ligation of these lattices with T4 DNA ligase should produce long covalent DNA strands. The nicks, where strands from adjacent DX units abut, are all on the upper or lower surface of the lattice, where they are accessible to the enzyme. The long strands can serve as reporters of successful lattice formation; shorter strands report either the presence of small aggregates or an occasional failure to ligate within the lattice. In the DAO-E design there are four vertical reporter strands, whereas in the DAE-O design there are two horizontal and two vertical reporter strands. All four reporter strands extend for more than 30 repeats when visualized by denaturing polyacrylamide gel electrophoresis (PAGE) (Fig. 2; data for DAO-E not shown). Longer strands co-migrate on this gel, so we cannot determine the full extent of polymerization. These results suggest that the lattice is a good substrate for T4 DNA ligase, and that the lattice can form with more than  $30 \times 30$  units. However, unintended associations or side

reactions could lead to similar distributions of strand lengths after ligation. Direct physical observation is necessary to confirm lattice assembly.

### AFM imaging

We have used atomic force microscopy<sup>16</sup> to demonstrate unequivocally the formation of 2-D lattices. **A** and **B** units are annealed separately, then combined and annealed together to form **AB** lattices. The resulting solution is deposited for adsorption on an atomically flat mica surface, and then imaged under isopropanol by contact mode AFM<sup>24</sup>. The solution is not treated with DNA ligase, and thus the lattices are held together only by noncovalent interactions (for example, hydrogen bonds and base stacking). This protocol ensures that the solution contains no protein contaminants and demonstrates that ligase activity is not necessary for the self-assembly process. Negative controls of buffer alone and of **A** or **B** alone show no aggregates larger than 20 nm (data not shown). In separate experiments, **A** and **B** DAO units were modified by the removal of two sticky ends from each unit (for example, all yellow and red sticky ends in Fig. 1d); when the modified **A** and **B** units were annealed together, we observed only linear and branched structures with apparent widths typically less than 10 nm (data not shown), providing additional negative controls. However, the unmodified **AB**



**Figure 2** Autoradiogram of a 4% denaturing polyacrylamide gel showing the product after assembly of 2-D lattice DAE-O (**AB**), ligation to form long covalent strands, and denaturing to separate the strands. Lane 1 contains a ladder or markers at 100-base intervals. In the remaining lanes, a different strand from unit **B** is labelled: lanes 2 and 3 correspond to red strands (Fig. 1c, right), lanes 4 and 5 are green strands, lane 6 and 7 are blue strands, and lanes 8 and 9 are yellow strands. *Escherichia coli* exonucleases I and III were added to the material in the odd-numbered lanes to degrade single-stranded species, indicating that no circular products were formed. As red strands are lengths 46 and 48, in lane 2 we see bands at lengths  $94k$  and  $94k + 48$  for integers  $k$ ; green strands are lengths 31 and 21, so lane 4 has bands at  $52k$  and  $52k + 21$ ; blue strands are both length 47, so lane 6 has bands at  $94k$  and  $94k + 47$ ; and yellow strands are lengths 21 and 31, so lane 8 has bands at  $52k$  and  $52k + 31$ .



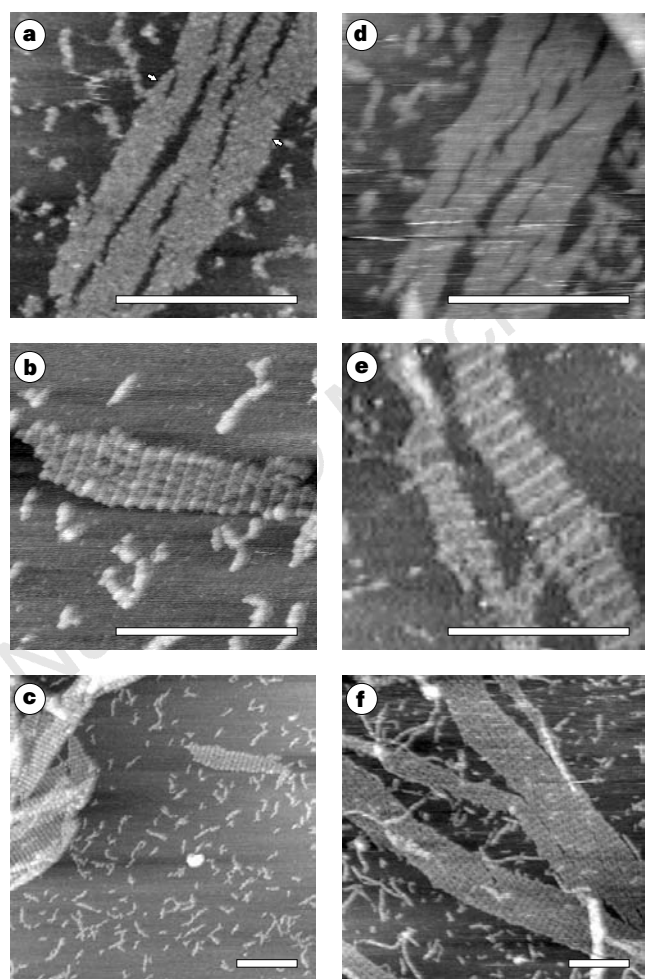
samples contain 2-D sheets many micrometres long, often more than 200 nm wide (Fig. 3a, d). The apparent height of the sheets is  $1.4 \pm 0.5$  nm, suggestive of a monolayer of DNA. The sheets often seem ripped and appear to have a grain, in that rips have a preferred direction consistent with the design (Fig. 1c). In the DAO-E lattice, a vertical rip requires breaking six sticky-end bonds per 12 nm torn, whereas a horizontal rip requires breaking only one sticky-end bond per 13 nm torn. A possible vertical column, perpendicular to the rips, is indicated in Fig. 3a (arrows). Although in this image the columns are barely perceptible, Fourier analysis shows a peak at  $13 \pm 1$  nm, suggesting that observed columns are 1 DX wide. Periodic topographic features would not be expected in the ideal AB lattice; however, a vertically stretched lattice may have gaps between the DX units that could produce the periodic features seen here. Because crystals are found in AFM samples taken from both the top and the bottom of the solution, we believe that crystals form in solution and are not due to interaction with a surface.

### Control of surface topography

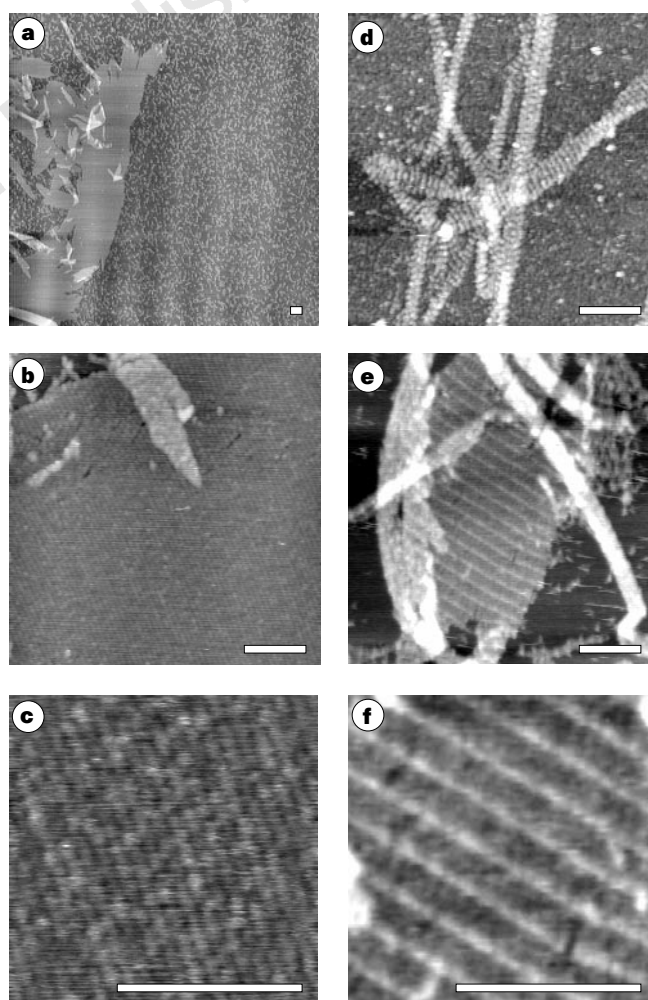
The self-assembling AB lattice can serve as scaffolding for other

molecular structures. We have decorated B with two DNA hairpin sequences inserted into its component strands, which we call B̂ (Fig. 1d). So decorated, the vertical columns of the lattice become strikingly apparent as stripes in AFM images (Fig. 3b, c, e, f), further confirming the proper self-assembly of the 2-D lattice. The spacing of the decorated columns is  $25 \pm 2$  nm for the DAO-E lattice and  $33 \pm 3$  nm for the DAE-O lattice, indicating that every other column is decorated, in accord with the design. Slow annealing at 20 °C and gentle handling of the DAO-E sample during deposition and washing has produced single crystals measuring up to  $2 \times 8 \mu\text{m}$  (Fig. 4a–c). Close examination shows that the stripes are continuous across the crystal, and thus it appears to be a single domain containing over 500,000 DX units.

Instead of DNA hairpins, other chemical groups can be used to label the DX molecules. Previously, biotin–streptavidin–gold has been used to label linear DNA for imaging by AFM<sup>25,26</sup>. We have used 1.4-nm nanogold–streptavidin conjugates to label DAE molecules. For these experiments, the central strand of B contains a 5' biotin group; after assembly of AB lattices, the solution containing DNA lattices is incubated with streptavidin–nanogold conjugates and



**Figure 3** AFM images of two-unit lattices. **a**, DAO-E AB lattice. A possible vertical column is indicated by the arrows. Fourier analysis shows  $13 \pm 1$  nm periodicity; each DAO is 12.6 nm wide. **b** and **c**, DAO-E AB̂ lattice (two views of the same sample). Stripes have  $25 \pm 2$  nm periodicity; the expected value is 25.2 nm. **d**, DAE-O AB lattice. **e** and **f**, DAE-O AB̂ lattice (for different samples, see Methods). Stripes have  $33 \pm 3$  nm periodicity; the expected value is 32 nm. All scale bars are 300 nm; images show  $500 \times 500$  nm or  $1.5 \times 1.5 \mu\text{m}$ . The grey scale indicates the height above the mica surface; the apparent lattice height is between 1 and 2 nm.



**Figure 4** AFM images showing large crystals and modifications of lattice periodicity and surface features. **a–c**, DAO-E AB̂ lattice at three levels of detail (all the same sample). The largest domain is  $\sim 2 \times 8 \mu\text{m}$ , and contains  $\sim 500,000$  DX units. **d**, DAE-O AB lattice in which B has been labelled with biotin–streptavidin–nanogold. **e** and **f**, DAE-O ABCD lattice at two levels of detail (the same sample). Stripes have  $66 \pm 5$  nm periodicity; the expected value is 64 nm. All scale bars are 300 nm; images show  $500 \times 500$  nm,  $1.5 \times 1.5 \mu\text{m}$ , or  $10 \times 10 \mu\text{m}$ . The grey scale indicates the height above the mica surface; the apparent lattice height is between 1 and 2 nm.

then imaged by AFM (Fig. 4d). We have not confirmed that the nanogold particles remain conjugated to the streptavidin molecules; the surface topography may be due to streptavidin molecules alone.

We have also tested DAO and DAE systems incorporating only one of the two hairpins in **B**, DAO systems in which the 3-arm junctions are relocated by two nucleotides towards the centre of the molecule, and DAE systems in which the arms are one nucleotide longer or shorter. All systems produced results similar to those shown in Fig. 3 when imaged by AFM (data not shown). The lattice assembly appears to be robust to variations in the local DX structure and is not sensitive to small variations in the annealing protocol. (The results reported here were obtained in two laboratories using different buffers, annealing conditions, and AFM instruments.)

Choice of sticky ends predictably determines the associations between DX units. Therefore it is straightforward to modify the **AB** system into a 4-DX system with twice the period (Fig. 1a, right). This requires the use of twice as many unique sticky-end sequences to control the unique association of DX units **A**, **B**, **C**, and **D**. We have created such a system in the DAE-O topology, where a single unit, **D**, is decorated with two hairpins. Crystals in this system show stripes spaced every  $66 \pm 5$  nm, confirming that every fourth vertical column is decorated (Fig. 4e, f).

In all images of **AB**, **AB** and **ABCD** systems, we observed many DNA structures in addition to the isolated 2-D crystals discussed above. In many images, the 2-D crystals appear to overlap, leading to discrete steps in thickness (Figs 3c, 4a, d). The arrangement of crystals on the mica—solitary, overlapping, piled up like driftwood, ripped to shreds—depends sensitively upon DNA concentration and upon the sample preparation procedure, especially the wash step. Prominently, the background of every image contains small objects, which we assume to be associations of small numbers of DX units. Also, long, thin ‘rods’ appear in some preparations (Figs 3f, 4d, lower right). These structures have not been characterized.

## Applications

Self-assembly is increasingly becoming recognized as a route to nanotechnology<sup>27</sup>. Our results demonstrate the potential of using DNA to create self-assembling periodic nanostructures. The periodic blocks used here are composed of either two or four individual DX units. However, the number of component tiles in the repeat unit does not appear to be limited to such small numbers, suggesting that complex patterns could be assembled into periodic arrays. These patterns could be either direct targets in nanofabrication or aids to the construction of such targets. Because oligonucleotide synthesis can readily incorporate modified bases at arbitrary positions, it should be possible to control the structure within the periodic block by decoration with chemical groups, catalysts, enzymes and other proteins<sup>28</sup>, metallic nanoclusters<sup>29,30</sup>, conducting silver clusters<sup>31</sup>, DNA enzymes<sup>32</sup>, or other DNA nanostructures such as polyhedra<sup>33,34</sup>.

It may be possible to extend the two-dimensional lattices demonstrated here into three dimensions. Designed crystals could serve as scaffolds for the crystallization of macromolecules<sup>2</sup>, as photonic materials with novel properties<sup>35</sup>, as designable zeolite-like materials for use as catalysts or as molecular sieves<sup>36</sup>, and as scaffolds for the assembly of molecular electronic components<sup>37</sup> or biochips<sup>38</sup>.

The self-assembly of aperiodic structures should also be considered. It may be possible to design molecular Wang tiles that self-assemble into aperiodic crystals according to algorithmic rules<sup>10,14</sup>. It will be crucial to understand the mechanisms of crystallogenesis and crystal growth in this system to provide a firm underpinning for the theoretical proposals of computation by self-assembly.

Progress in this field will require detailed knowledge of the physical, kinetic, structural, dynamic and thermodynamic parameters that characterize DNA self-assembly. Additionally, improved methods for error reduction and purification must be developed.

The approach described here provides a uniquely versatile experimental system for investigating these issues. □

## Methods

**DNA sequences and synthesis.** Figure 1d shows sequences used in these experiments; for historical reasons, some figures show experiments where variants of these sequences were used. The sequences for DAO-E **A**, **B**, and **B** given in Fig. 1d were used for Fig. 3b, c. Figures 3a, 4a–c, show DAO systems with symmetrical sticky ends: the sequence for the green strands of DAO **A** is 5'TCACT...GAGAT3' and the sequence for the blue strands of DAO **B** and **B** is 5'AGTGA...ATCTC3'. The sequences for the DAE-O **A** and **B** units used in Fig. 3d are modified from those given in Fig. 1d by deletion of the rightmost A:T base pair in the upper helix of **A**, addition of C to the upper-rightmost 5' end of **A**, and deletion of C from the lower-leftmost 3' end of **B**; also the central strand of **A** is cyclically permuted to begin 5'GCATGT... and the central strand of **B** is cyclically permuted to begin 5'GGTCAC.... For Fig. 3e, the sequences for the DAE-O **A** are as shown in Fig. 1d, but both hairpins of **B** stem from the lower helix; thus the upper strand is as in **B**, and the central strand is 5'ACTGGTT AGTGGATTGCGTAGGCGAGTAGTTTTCTACTCGCTTTACAACGCCACCG ATGCGGTC3'. The sequences for the DAE-O **A** and **B** units used in Fig. 3f are exactly as shown in Fig. 1d. Sequences for **ABCD** are available as Supplementary Information. All oligonucleotides were synthesized by standard methods, PAGE purified, and quantitated by ultraviolet absorption at 260 nm in H<sub>2</sub>O.

**Annealing of oligonucleotides.** The strands of each DX unit were mixed stoichiometrically and dissolved to 0.2–2  $\mu$ M in TAE/Mg<sup>2+</sup> buffer (40 mM Tris-HCl (pH 8.0), 1 mM EDTA, 3 mM Na<sup>+</sup>, 12.5 mM Mg<sup>2+</sup>) or in a HEPES buffer (10 mM HEPES, 6 mM MgCl<sub>2</sub>, 1 mM EDTA, pH 7.8). The solutions were annealed from 90 °C to room temperature over the course of several hours in a Perkin-Elmer polymerase chain reaction (PCR) machine (to prevent concentration by evaporation) or were annealed from 100 °C to room temperature during 40 h in a 2-litre water bath insulated in a styrofoam box. To produce lattices, equal amounts of each DX were mixed and annealed from 50 °C to 20 °C for up to 36 h. In some cases (Fig. 3a–c), all strands were mixed together from the start.

**Gel electrophoresis.** For gel-based studies, T4 polynucleotide kinase (Amersham) was used to phosphorylate strands with <sup>32</sup>P; these strands were then PAGE-purified and mixed with an excess of unlabelled strands. Non-denaturing 4, 5, or 8% PAGE (19:1 acrylamide:bisacrylamide) in TAE/Mg<sup>2+</sup> was done at 4 °C or at room temperature. For denaturing experiments, after annealing in T4 DNA ligase buffer (Amersham) (66 mM Tris-HCl (pH 7.6), 6.6 mM MgCl<sub>2</sub>, 10 mM DTT, 66  $\mu$ M ATP), 1  $\mu$ l containing 10 units of T4 DNA ligase (Amersham) was added to 10  $\mu$ l DNA solution and incubated for up to 24 h at 16 °C or at room temperature. For exonuclease reactions, 50 units of exonuclease III (Amersham) and 5 units of exonuclease I (Amersham) were added after ligation and incubated an additional 3.5 h at 37 °C. The solution was added to an excess of denaturing dye buffer (0.1% xylene cyanol FF tracking dye in 90% formamide with 1 mM EDTA, 10 mM NaOH) and heated to 90 °C for at least 5 min before loading. Denaturing gels contained 4% acrylamide (90:1 acrylamide:bisacrylamide) and 8.3 M urea in TBE (89 mM Tris-HCl (pH 8.0), 89 mM boric acid, 2 mM EDTA). Gels were analysed by using a Phosphorimager.

**Labelling with biotin-streptavidin-nanogold.** The central strand of DAE-O **B** was synthesized containing a 5' biotin group. Lattices were formed by annealing as described. After annealing, a stoichiometric amount of streptavidin-nanogold (1.4 nm; Nanoprobe) in a buffer provided by the supplier (20 mM phosphate, 150 mM NaCl (pH 7.4), 0.1% BSA, 0.05% NaN<sub>3</sub>) was added, so that the molar ratio of biotin to streptavidin was 1:1. The solution was left at room temperature for 1 h and then imaged by AFM as described below.

**Preparation of AFM sample.** 2–10  $\mu$ l were spotted onto freshly cleaved mica (Ted Pella) and left to adsorb to the surface for 2 min. To remove buffer salts, 5–10 drops of doubly distilled or nanopure H<sub>2</sub>O were placed on the mica, the drop was shaken off and the sample was dried with compressed air. Imaging was done under isopropanol in a fluid cell on a NanoScope II using the D or E scanner and commercial 200- $\mu$ m cantilevers with Si<sub>3</sub>N<sub>4</sub> tips (Digital Instruments). The feedback setpoint was adjusted frequently to minimize contact force to ~1 to 5 nN. Images were processed with a first- or third-order ‘flatten filter’, which independently subtracts a first- or third-order polynomial

fit from each scanline to remove tip artefacts; however, this technique can introduce false 'shadows'.

Received 10 March; accepted 5 May 1998.

1. Liu, B., Leontis, N. B. & Seeman, N. C. Bulged 3-arm DNA branched junctions as components for nanoconstruction. *Nanobiology* **3**, 177–188 (1994).
2. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–247 (1982).
3. Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids *in vitro*. *Proc. Natl Acad. Sci. USA* **70**, 3240–3244 (1973).
4. Qiu, H., Dewan, J. & Seeman, N. C. A DNA decamer with a sticky end: The crystal structure of d-CGACGATCGT. *J. Mol. Biol.* **267**, 881–898 (1997).
5. Ma, R.-L., Kallenbach, N. R., Sheardy, R. D., Petrillo, M. L. & Seeman, N. C. Three-arm nucleic acid junctions are flexible. *Nucleic Acids Res.* **14**, 9745–9753 (1986).
6. Petrillo, M. L. *et al.* The ligation and flexibility of four-arm DNA junctions. *Biopolymers* **27**, 1337–1352 (1988).
7. Fu, T.-J. & Seeman, N. C. DNA double-crossover molecules. *Biochemistry* **32**, 3211–3220 (1993).
8. Schwacha, A. & Kleckner, N. Identification of double Holliday junctions as intermediates in meiotic recombination. *Cell* **83**, 783–791 (1995).
9. Li, X., Yang, X., Qi, J. & Seeman, N. C. Antiparallel DNA double crossover molecules as components for nanoconstruction. *J. Am. Chem. Soc.* **118**, 6131–6140 (1996).
10. Winfree, E. in *DNA Based Computers: Proceedings of a DIMACS Workshop, April 4, 1995, Princeton University* (eds Lipton, R. J. & Baum, E. B.) 199–221 (American Mathematical Society, Providence, RI, 1996).
11. Adleman, L. M. Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994).
12. Grünbaum, B. & Shephard, G. C. *Tilings and Patterns* (Freeman, New York, 1986).
13. Wang, H. in *Proc. Symp. Math. Theory of Automata* 23–56 (Polytechnic, New York, 1963).
14. Winfree, E., Yang, X. & Seeman, N. C. in *Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University, June 20–12, 1996* (American Mathematical Society, Providence, RI, in the press).
15. Reif, J. in *Proceedings of the 3rd DIMACS Meeting on DNA Based Computers, University of Pennsylvania, June 23–25, 1997* (American Mathematical Society, Providence, RI, in the press).
16. Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
17. Wang, J. C. Helical repeat of DNA in solution. *Proc. Natl Acad. Sci. USA* **76**, 200–203 (1979).
18. Rhodes, D. & Klug, A. Helical periodicity of DNA determined by enzyme digestion. *Nature* **286**, 573–578 (1980).
19. Seeman, N. C. *De novo* design of sequences for nucleic acid structural engineering. *J. Biomol. Struct. Dyn.* **8**, 573–581 (1990).
20. Yue, K. & Dill, K. A. Inverse protein folding problem—designing polymer sequences. *Proc. Natl Acad. Sci. USA* **89**, 4163–4167 (1992).
21. Sun, S., Brem, R., Chan, H. S. & Dill, K. A. Designing amino acid sequences to fold with good hydrophobic cores. *Prot. Engng* **9**, 1205–1213 (1996).
22. SantaLucia, J., Allawi, H. T. & Seneviratne, A. Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* **35**, 3555–3562 (1996).
23. Ouporov, I. V. & Leontis, N. B. Refinement of the solution structure of a branched DNA three-way junction. *Biophys. J.* **68**, 266–274 (1995).

24. Hansma, H. G. *et al.* Reproducible imaging and dissection of plasmid DNA under liquid with the atomic force microscope. *Science* **256**, 1180–1184 (1992).
25. Shaiu, W.-L., Larson, D. D., Vesenska, J. & Henderson, E. Atomic force microscopy of oriented linear DNA molecules labelled with 5 nm gold spheres. *Nucleic Acids Res.* **21**, 99–103 (1993).
26. Shaiu, W.-L., Vesenska, J., Jondle, D., Henderson, E. & Larson, D. D. Visualization of circular DNA molecules labelled with colloidal gold spheres using atomic force microscopy. *J. Vac. Sci. Tech. A* **11**, 820–823 (1993).
27. Whitesides, G. M., Mathias, J. P. & Seto, C. T. Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures. *Science* **254**, 1312–1319 (1991).
28. Niemeyer, C. M., Sano, T., Smith, C. L. & Cantor, C. R. Oligonucleotide-directed self-assembly of proteins. *Nucleic Acids Res.* **22**, 5530–5539 (1994).
29. Alivisatos, A. P. *et al.* Organization of 'nanocrystal molecules' using DNA. *Nature* **382**, 609–611 (1996).
30. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).
31. Braun, E., Eichen, Y., Sivan, U. & Ben-Yoseph, G. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature* **391**, 775–778 (1998).
32. Breaker, R. R. & Joyce, G. F. A DNA enzyme that cleaves RNA. *Chem. Biol.* **1**, 223–229 (1994).
33. Chen, J. & Seeman, N. C. The synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–633 (1991).
34. Zhang, Y. & Seeman, N. C. The construction of a DNA truncated octahedron. *J. Am. Chem. Soc.* **116**, 1661–1669 (1994).
35. Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals: Moulding the Flow of Light* (Princeton University Press, Princeton, 1995).
36. Ribeiro, F. R. *et al.* Structure–activity relationships in zeolites. *J. Mol. Cat. A: Chem.* **96**, 245–270 (1996).
37. Robinson, B. H. & Seeman, N. C. The design of a biochip: A self-assembling molecular-scale memory device. *Prot. Engng* **1**, 295–300 (1987).
38. Haddon, R. C. & Lamola, A. A. The molecular electronic device and the biochip computer: present status. *Proc. Natl Acad. Sci. USA* **82**, 1874–1878 (1985).
39. Carter, E. S. & Tung, C.-S. NAMOT2—a redesigned nucleic acid modelling tool: construction of non-canonical DNA structures. *CABIOS* **12**, 25–30 (1996).
40. Vainshtein, B. K. *Modern Crystallography, 1: Fundamentals of Crystals* (Springer, New York, 1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank J. Hopfield, S. Rowles, S. Mahajan, C. Brody, L. Adleman and P. Rothmund for discussion; J. Abelson and his group for use of his laboratory and for technical advice; A. Segal, E. Rabani and R. Moision for instruction and advice on AFM imaging; the Beckman Institute Molecular Materials Resource Center for assistance and use of their AFM facilities; F. Furuya for help with labelling; and M. Yoder, V. Morozov, D. Stokes, M. Simon and J. Wall for assistance in early attempts to visualize DNA lattices. The research at Caltech has been supported by the National Institute for Mental Health, General Motors' Technology Research Partnerships program, and by the Center for Neuro-morphic Systems Engineering as a part of the NSF Engineering Research Center Program. The research at NYU has been supported by the Office of Naval Research, the National Institute of General Medical Sciences, and the NSF.

Correspondence and requests for materials should be addressed to E.W. (e-mail: winfree@hope.caltech.edu) or N.C.S. (e-mail: ned.seeman@nyu.edu).

# KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited  
90 Tottenham Court Road, London W1P 0LP  
Telephone: 0171 436 5931 Fax: 0171 436 3986



# Pitx2 determines left-right asymmetry of internal organs in vertebrates

Aimee K. Ryan<sup>\*,†</sup>, Bruce Blumberg<sup>†,‡</sup>, Concepción Rodríguez-Esteban<sup>†,‡</sup>, Sayuri Yonei-Tamura<sup>†,‡</sup>, Koji Tamura<sup>‡</sup>, Tohru Tsukui<sup>‡</sup>, Jennifer de la Peña<sup>‡</sup>, Walid Sabbagh<sup>‡</sup>, Jason Greenwald<sup>‡</sup>, Senyon Choe<sup>‡</sup>, Dominic P. Norris<sup>§</sup>, Elizabeth J. Robertson<sup>§</sup>, Ronald M. Evans<sup>‡,||</sup>, Michael G. Rosenfeld<sup>\*</sup> & Juan Carlos Izpisua Belmonte<sup>‡</sup>

<sup>\*</sup> Howard Hughes Medical Institute, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0648, USA

<sup>§</sup> Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA

<sup>||</sup> Howard Hughes Medical Institute, <sup>‡</sup> The Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

<sup>†</sup> These authors contributed equally to this work

**The handedness of visceral organs is conserved among vertebrates and is regulated by asymmetric signals relayed by molecules such as Shh, Nodal and activin. The gene *Pitx2* is expressed in the left lateral plate mesoderm and, subsequently, in the left heart and gut of mouse, chick and *Xenopus* embryos. Misexpression of Shh and Nodal induces *Pitx2* expression, whereas inhibition of activin signalling blocks it. Misexpression of *Pitx2* alters the relative position of organs and the direction of body rotation in chick and *Xenopus* embryos. Changes in *Pitx2* expression are evident in mouse mutants with laterality defects. Thus, *Pitx2* seems to serve as a critical downstream transcription target that mediates left-right asymmetry in vertebrates.**

The vertebrate body exhibits bilateral symmetry externally whereas the internal organs display significant left-right asymmetry. During organogenesis, the unpaired organs of the chest and abdomen begin development in the midline and then lateralize, with the first morphological markers of left-right asymmetry being the right-sided looping of the developing heart. A second sign of asymmetry is then manifested by the rotation of the body in amniote embryos. Virtually all visceral organs ultimately show left-right asymmetry, either with respect to their location in the body cavity or by morphological differences on one side versus the other. The left-right asymmetries of internal organ placement are invariant within a given species and have been conserved throughout evolution. Normal organ placement is termed *situs solitus*, and the mirror-image arrangement is *situs inversus*. Other defects of *situs* are partial (heterotaxy) or complete (isomerism) loss of asymmetry. Left-right axis malformations in humans are phenotypically variable and genetically heterogeneous<sup>1,2</sup>. Generally, individuals with complete *situs inversus* do not suffer severe clinical consequences, whereas heterotaxia and isomerism are associated with moderate-to-severe physiological complications<sup>3,4</sup>.

As the establishment of correct left-right asymmetry is critical for survival, the mechanisms governing initiation and maintenance of these asymmetries should be tightly regulated and evolutionarily conserved. Several models have been proposed to account for these asymmetries (reviewed in refs 5–7). In chick, there is a signalling cascade involving members of the TGF- $\beta$  superfamily, namely activin- $\beta$ B and Nodal, the activin receptor RIIA (cAct-RIIA) and Sonic hedgehog (Shh), all of which are asymmetrically expressed with respect to the left-right axis<sup>8,9</sup>. *Activin- $\beta$ B*, present asymmetrically on the right side of stage 3–5+ embryos<sup>9,10</sup>, is thought to induce local expression of *cAct-RIIA*<sup>8,10</sup>, which in turn represses the bilaterally symmetrical *Shh* expression in Hensen's node on the right<sup>8,9</sup>. This leads to left-sided expression of *Shh* and induction of *nodal* in the left lateral plate mesoderm<sup>8</sup>. Misexpression of activin or Shh disrupts the normal expression pattern of *nodal* and randomizes heart looping. In *Xenopus*, inappropriate expression of the TGF- $\beta$  family member Vg-1 inverts *nodal* expression and results in *situs inversus*<sup>11,12</sup>. In contrast to the chicken model, targeted gene

deletion of Shh, activin- $\beta$ B, follistatin or Act-RIIA in mice does not alter the left-right orientation of the heart or of the internal organs, calling into question their role in left-right patterning in the mouse<sup>13–17</sup>. Mice null for *Act-RIIB*, which is not asymmetrically expressed in chick or mouse, exhibit defects in left-right asymmetries, including isomerisms<sup>18</sup>, suggesting that Act-RIIB is a critical component of the left-right pathway in mouse.

Of the many molecules that have been implicated in left-right signalling during vertebrate embryogenesis, only Nodal exhibits a clear correlation between its expression in the lateral plate mesoderm and visceral *situs*<sup>19,20</sup>. In *inv/inv* mice, where virtually all animals exhibit *situs inversus*, *nodal* is expressed only in the right lateral plate mesoderm<sup>19,20</sup>. In *iv* mice, where left-right development is randomized, all four possible patterns of *nodal* expression are observed: left, right, bilateral and absent<sup>20</sup> (see also ref. 21). *nodal* expression is bilateral in *Fused toes*<sup>22</sup> and *no turning*<sup>23</sup> mice, which also have randomized left-right asymmetries. Altering the normal *nodal* expression pattern in the left lateral plate mesoderm in *Xenopus* and chick is also associated with changes in left-right development<sup>8,11,24–26</sup>. Thus, Nodal appears to be a conserved factor in the cascade that establishes left-right asymmetry in all vertebrates. The observations that *nodal* expression reliably predicts *situs* and that loss of Act-RIIB function leads to defects in *situs* suggests that these factors function in a common signalling pathway.

Although progress has been made in understanding early events in the determination of left-right asymmetry, much is yet to be learned about how multiple extracellular signals are transduced, propagated and maintained, ultimately leading to visceral asymmetry. Transcription factors are good candidates for mediating these processes. However, relatively little is known of their role in this process, and only three have been implicated in the left-right asymmetry pathway. HNF-3 $\beta$  may have a role because it is transiently asymmetrically expressed in the chick<sup>8</sup> and because *HNF-3 $\beta$ <sup>+/-</sup>*, *nodal*<sup>lacZ/+</sup> double-heterozygous mice express *lacZ* bilaterally in the lateral plate mesoderm and have defects in the positioning of the viscera and heart, and random embryonic rotation<sup>19</sup>. The zinc-finger gene *Snail-Related* (*cSnR*) which is initially expressed

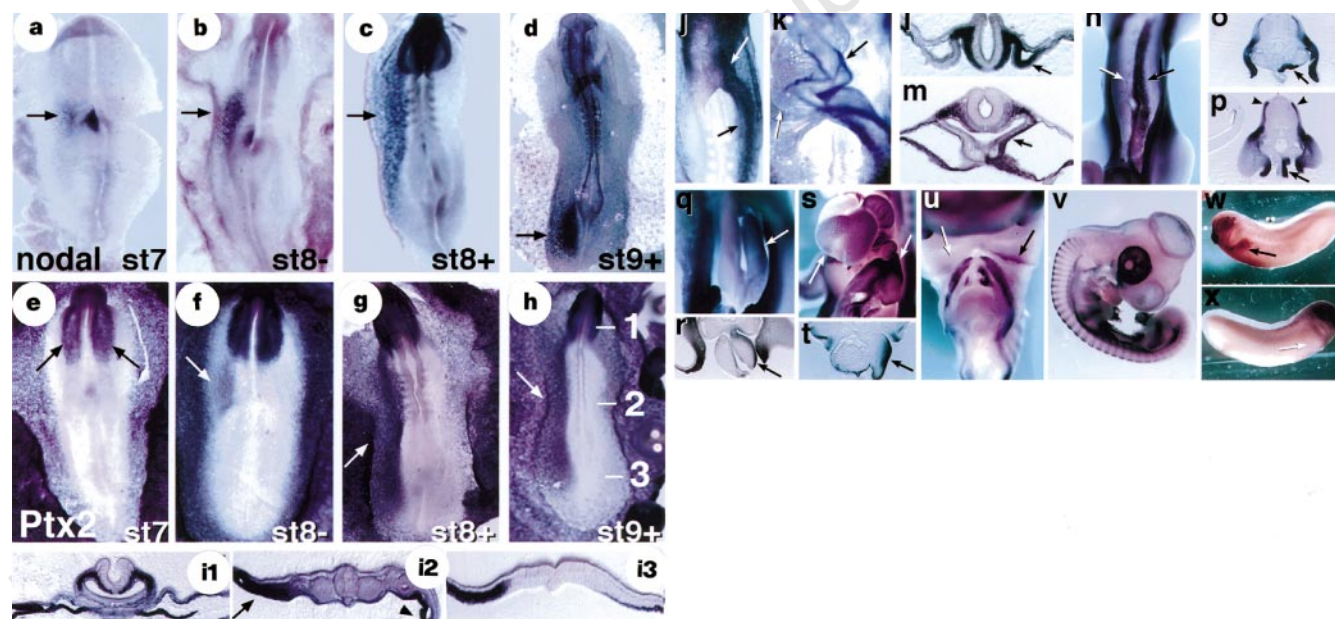
bilaterally in the presumptive anterior cardiac mesoderm before becoming significantly more intense on the right, is downregulated by ectopic expression of *Shh* on the right, and perturbed by ectopic activin on the left. Antisense experiments designed to disrupt *cSnR* translation reverse heart looping<sup>27</sup>. Finally, the homeodomain factor Nkx2.5 appears to regulate the asymmetric expression of the basic helix–loop–helix (bHLH) factors dHAND and eHAND, which are required for correct heart looping and morphogenesis<sup>28,29</sup>.

Here we investigate the role of the bicoid-related homeodomain transcription factor Pitx2 in determining left–right asymmetry in chick, *Xenopus* and mouse. The human homologue of Pitx2, RIEG, was originally described as the gene for Rieger syndrome<sup>30</sup>, an autosomal dominant human disorder characterized by ocular anterior chamber anomalies, dental hypoplasia, mild craniofacial dysmorphism and umbilical stump abnormalities, together with occasional defects in cardiac, limb and pituitary development. Our results indicate that Pitx2 may turn on the gene network responsible for the morphological events that result in left–right asymmetries in vertebrates. Whereas umbilical and cardiac phenotypes may suggest a link between Pitx2 and heart and gut development, the lack of alteration in organ *situs* in individuals affected with Rieger syndrome may be due to the presence of the wild-type allele. In chick, *Xenopus* and mouse, *Pitx2* expression is on the left side of the

embryo in the lateral plate mesoderm; it then continues to be expressed asymmetrically in several organs that are asymmetric with respect to the left–right axis of the embryo. *Pitx2* expression in the left lateral plate mesoderm is preceded by *Shh* and *nodal*, and we find that *Pitx2* expression can be induced by both *Shh* and *Nodal*, suggesting that it is downstream of these signalling molecules. In mutant mice with laterality defects, *Pitx2* expression correlates with changes of visceral *situs*, paralleling the expression of *nodal*. Inhibition of signalling through a dominant-negative activin type-II receptor also alters *Pitx2* expression. Finally, ectopic expression of *Pitx2* in the right lateral plate mesoderm results in isomerism, or in reversed looping of the heart and gut and reversed body rotation in chick and *Xenopus* embryos. Our results indicate that Pitx2 may interpret and subsequently execute the left–right developmental program dictated by upstream signalling molecules and they identify Pitx2 as the first evolutionarily conserved transcription factor in the left–right pathway to control embryonic handedness in vertebrates.

### Asymmetric expression during embryogenesis

Chick and *Xenopus* *Pitx2* (mammalian homologues *RIEG*<sup>30</sup>, *Pitx2* (ref. 31), *Potx2* (ref. 32), and *Apr-1* (ref. 33)) were isolated by screening chick and *Xenopus* complementary DNA libraries with the murine P-OTX cDNA clone<sup>34,35</sup>. The overall identities for these *Pitx2*



**Figure 1** *Pitx2* is expressed asymmetrically in the lateral plate mesoderm, heart and gut. In this and other figures, *Pitx2* is referred to by its old name *Ptx2*. **a–i**, Dorsal views; **j–x**, ventral views. Where views are dorsal, the left side of the embryo corresponds to the reader's left; for ventral views, the left side of the embryo corresponds to the reader's right. **a**, *Nodal* is expressed in a small medial domain immediately lateral and posterior to the node and in the adjacent region of the left lateral plate mesoderm (arrow). **b**, At stage 8 *nodal* expression in the left lateral plate mesoderm (arrow) extends posteriorly. **c**, *Nodal* is expressed throughout the left lateral plate mesoderm (arrow) at stage 8<sup>+</sup>. Expression of *nodal* is also seen in the head mesenchyme. **d**, At stage 9<sup>+</sup>, a strong region of *nodal* expression is detected in the posterior region of the left lateral plate mesoderm. **e**, Stage-7 embryo, showing the symmetrical expression of *Pitx2* in the head mesenchyme (arrows). **f, g**, Dorsal views of stage 8<sup>+</sup> to 9<sup>+</sup> embryos showing *Pitx2* expression in the head mesenchyme and the left lateral plate mesoderm (arrows). **h, i**, Transverse sections (**i**) through stage 9<sup>+</sup> embryo shown in **h**: the level at which the sections 1, 2 and 3 were obtained is indicated in **h**. **i**(2), *Pitx2* expression is symmetrically detected in the head mesenchyme; in **i**(2), *Pitx2* expression is detected in both the lateral plate mesoderm and endoderm on the left (arrow) but only in the endoderm on the right (arrowhead). **j**, Stage 10 embryo. *Pitx2* is expressed on the left side of the heart tube (white arrow) and left vitelline

vein (black arrow). **k**, Stage 12 embryo. *Pitx2* is expressed on the left (black arrow), but not on the right (white arrow) side of the looping heart tube. **l, m**, Transverse section of the embryos shown in **j** and **k**, respectively, showing the expression of *Pitx2* in the left heart tube before looping (arrow in **l**) and in the left side of the epimyocardium (arrow in **m**) during heart looping. **n**, Stage 19 embryo. *Pitx2* expression is present on the left side (white arrow) but not the right side (black arrow) of the developing gut and crop. **o, p**, Transverse sections of the same embryo showing asymmetric *Pitx2* expression on the left side of the developing gut (arrows). Bilateral expression is observed in the somites (arrowheads) and at the proximal aspects of the developing hindlimb in **p**. **q**, Stage 21 embryo. *Pitx2* mRNA is detected in the left portion of the caeca (arrow) but not on the right. This is also seen in the transverse section in **r** (arrow). **s**, Ventral view of stage 25 embryo. *Pitx2* is expressed in the heart ventricle (left arrow) and in the gizzard (right arrow). **u**, Stage 22 embryo, showing stronger *Pitx2* expression on the left of the second branchial arch (black arrow). **v**, Stage 24 embryo with *Pitx2* expression in the eye, somites and limb muscle. **w, x**, In stage 26 *Xenopus* embryos, *Pitx2* is expressed bilaterally in the eye and cement gland. Expression of *Pitx2* occurs in the left lateral plate mesoderm (**w**, arrow) but not in the right lateral plate mesoderm (**x**, arrow).

homologues are: 98% mouse versus human, 97% chicken versus mouse, 96% chicken versus human, and 89% *Xenopus* versus chicken, mouse or human (see Supplementary Information for amino-acid sequence alignment).

Whole-mount *in situ* hybridization was used to reveal the temporal and spatial expression of *Pitx2* messenger RNA during chick embryogenesis. *Pitx2* mRNA is first detected in the head mesenchyme without apparent left–right asymmetry at stage 7 (Fig. 1e). At stage 8<sup>+</sup>, *Pitx2* expression is maintained in the head mesenchyme and appears in a small region of the left lateral plate mesoderm (Fig. 1f). By stage 8<sup>+</sup>, *Pitx2* mRNA can be detected along the entire left side of the lateral plate mesoderm (Fig. 1g) and this expression pattern remains relatively unchanged through to stage 10 (Fig. 1h, i). The onset of *Pitx2* expression in the left lateral plate mesoderm appears to be later than that of *Nodal*, which is also expressed in the left lateral plate mesoderm between stages 7–10 (Fig. 1a–d). At stage 10 *Pitx2* is expressed in the developing left heart tube, and later on in the left heart once it starts to loop (Fig. 1j–m, and data not shown). At stage 22, expression in the second branchial arch is stronger on the left side (Fig. 1u). Asymmetric *Pitx2* expression is also observed in the developing gut, caeca, gizzard and intestine between stages 19 and 25 (Fig. 1n–t). Bilaterally symmetrical expression of *Pitx2* in the somites is first apparent at stage 19 and intensifies as development proceeds (Fig. 1p, v). By stage 22, *Pitx2* is expressed in the mesenchymal cells that migrate into the limb buds and give rise to the limb muscles (Fig. 1v, and data not shown).

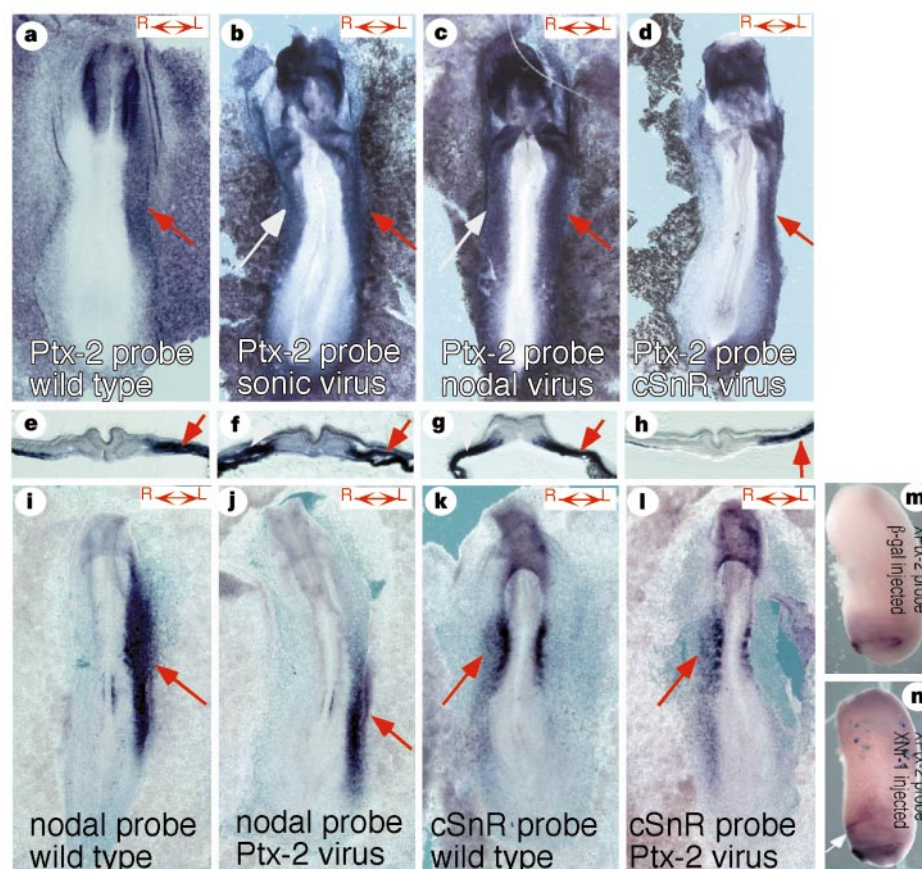
The spatial expression of *Pitx2* mRNA in *Xenopus* and mouse was also investigated using whole-mount *in situ* hybridization. In *Xenopus*, asymmetric expression of *Pitx2* is observed first at stage 24, becoming more intense as development proceeds in the left lateral plate mesoderm (Fig. 1w, x). At later stages, and as in the

chick, *Pitx2* is expressed on the left half of the heart tube at stage 30 and on the left side of the developing intestine at stage 42 (data not shown). A similar asymmetric pattern in the lateral plate mesoderm is seen in the mouse embryo (see Fig. 4). During heart development, *Pitx2* transcripts are detected in the left half of the linear heart tube (six somite embryo) and later on, at heart looping, in the left side of the ventricular, outflow tract and atrium. During the looping of the gastrointestinal tract, *Pitx2* is expressed on the left side of the linear gut tube in the six-somite embryo and on the left side of the stomach (E11–E16) and caeca (E14–E16) (data not shown).

Two observations suggest a role for *Pitx2* in generating left–right asymmetries. First, the asymmetric expression of *Pitx2* in the left lateral plate mesoderm of chick, *Xenopus* and mouse is preceded by the appearance of *nodal* transcripts in the same region<sup>8,19,20,36</sup>. Second, in contrast to transient expression of *nodal*, *Pitx2* expression is maintained on the left side during organogenesis of the heart and the gastrointestinal tract. These results indicate that *Pitx2* expression may be regulated by signalling molecules previously shown to participate in determining left–right asymmetry, such as *Shh* and *Nodal*.

### Shh and Nodal induce *Pitx2*

To test this possibility, we ectopically expressed *Shh* and *nodal* in the right lateral plate mesoderm of stage-4–6 chick embryos by infection *in ovo* with a replication-competent avian retrovirus containing either *Shh* or *nodal* cDNAs. To assay the reliability of *in ovo* infection, a replication-competent retrovirus expressing green fluorescent protein (GFP) was injected to the right of the node at stage 4 and the extent of viral infection was visualized at stages 8–12. Although many embryos showed a variable level of infection (about half of the injected embryos were poorly infected), about 25% of the injected blastoderms showed strong GFP expression on the right side of the embryo (data not



**Figure 2** *Pitx2* is downstream of Sonic hedgehog and Nodal. Whole-mount *in situ* hybridization of chick (a–l) and *Xenopus* (m, n) embryos; ventral views. a, Normal *Pitx2* expression in the head mesenchyme and in the left lateral plate mesoderm (arrow). Embryos infected in the right lateral plate mesoderm with RCAS-*Shh* (b) or RCAS-*lital-nodal* (c) exhibit bilateral *Pitx2* expression in the lateral plate mesoderm (red and white arrows indicate endogenous and ectopic expression respectively). Expression of *Pitx2* is not altered in embryos infected with the zinc-finger protein *Snail-Related* (d). e–h, Transverse sections through the flank of embryos shown in a–d, respectively. White arrows indicate ectopic expression. (Red arrows indicate endogenous expression.) i, Expression of *nodal* in left lateral plate mesoderm of a wild-type embryo. j, Expression of *nodal* is unchanged in an embryo infected with RCAS-*Pitx2* (arrow). k, Expression of *cSnR* in the right lateral plate mesoderm of an uninfected embryo (arrow). l, Expression of *cSnR* is unaffected in embryos infected with RCAS-*Pitx2*. m, *Pitx2* is not expressed in right lateral plate mesoderm of a *Xenopus* embryo injected with CMV-βgal. n, *Pitx2* expression in the right lateral plate mesoderm of a *Xenopus* embryo injected with *Xnr-1* (white arrow). R → L in the right top portion of the panels indicates the orientation of the embryo, R being right and L being left.



shown). In addition, infection *in ovo* does not result in the artefactual changes in *situs* often seen *in vitro* (see below and ref. 8).

*Shh* and *nodal* injected embryos were fixed between stages 8–12 and expression of *Pitx2* was assessed using whole-mount *in situ* hybridization. Ectopic expression of *Shh* or *nodal* caused bilateral expression of *Pitx2* in the lateral plate mesoderm (Fig. 2a–c, e–g). We also tested the effect of misexpressing the zinc-finger gene *Snail-Related*, *cSnR*, which is required for correct sidedness of heart looping<sup>27</sup>. However, misexpression of *cSnR* in the early chick embryo did not affect the expression of *Pitx2* (Fig. 2d, h).

Similar experiments to determine the ability of Nodal to induce *Pitx2* expression were also done in *Xenopus*. Plasmids expressing *Xnr-1* under the control of the strong cytomegalovirus (CMV) promoter were microinjected into one blastomere of the 4-cell embryo and assayed for *Pitx2* expression at stage 24–26. 86% of the injected embryos that showed exclusive expression of the lineage tracer on the right side exhibited ectopic expression of *Pitx2* on the right side (Fig. 2m, n; see Methods for the number of embryos injected).

To investigate the genetic hierarchy between *Shh*, *nodal* and *Pitx2*, we misexpressed *Pitx2* in the right lateral plate chick mesoderm. There were no detectable changes in *Shh*, *nodal* or *cSnR* expression (Fig. 2i–l and data not shown). Microinjection of *Pitx2* mRNA into *Xenopus* embryos did not affect *Xnr-1* expression either (data not shown). These results suggest that *Pitx2* is downstream of *Shh* and *Nodal*, and perhaps in a parallel pathway to that of *cSnR-1*.

### ***Pitx2*, organ asymmetry and body rotation**

The asymmetric expression of *Pitx2* in the left lateral plate mesoderm, and the observation that its expression can be induced in right lateral plate mesoderm in response to ectopic expression of *Shh* and *nodal* in this region, indicate that *Pitx2* may be one of the downstream effectors in the signalling pathway leading to left–right asymmetry. We therefore infected the right lateral plate mesoderm in stage-4–6 chick embryos with an RCAS retroviral vector containing full-length *Pitx2* and scored embryos for changes in heart morphology at stages 11–13. About half of the injected embryos were truncated with respect to their anterior–posterior axis and had a relatively normal heart whose location was shifted towards the head region. The remaining embryos appeared grossly normal and ~55% of these embryos had defects in heart *situs* (Fig. 3). Most of the affected embryos (~70%) had a bilaterally symmetrical heart (isomerism) that was centrally located with respect to the left–right axis (Fig. 3a, b, e), about 25% had reversed heart looping (Fig. 3a, c, d, f), and occasionally double hearts were found (data not shown). Misexpression of GFP did not induce reversal of heart looping or heart isomerisms.

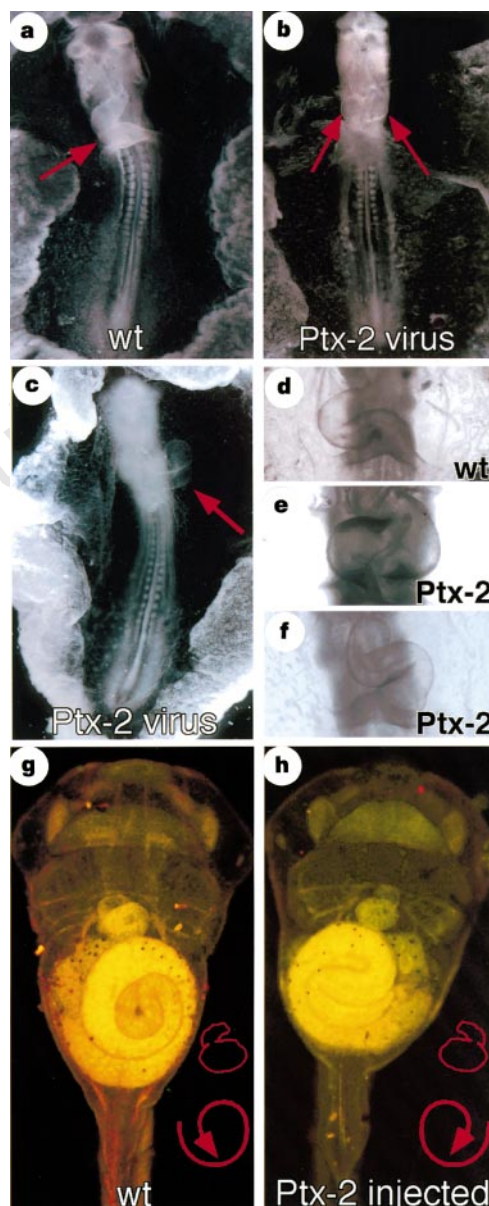
When embryos injected with *Pitx2* at stage 4 were left to develop further, about 12% of the embryos showed a reversal in the direction of embryonic rotation (the embryo turned to the left instead of turning to the right as normal; data not shown). This result suggests that the stronger expression of *Pitx2* on the left side of the body might be relevant at later stages in specifying the direction of body rotation. Similar results regarding the regulation of *Pitx2* by *nodal* and *Shh*, as well as its ability to induce heart isomerism and reversal of heart looping and body rotation, have been obtained independently<sup>37</sup>.

Microinjection of *Pitx2* mRNA into a two-cell *Xenopus* embryo showed that, as in the chicken, *Pitx2* overexpression causes alterations in left–right asymmetry (Fig. 3g, h). The predominant phenotype observed was heterotaxy, with the heart being more frequently reversed than the direction of gut coiling, although a few embryos with complete *situs inversus* were obtained (see Fig. 3h, for example).

### **Altered *Pitx2* expression in *iv* and *inv* mice**

Studies of *nodal* expression in *iv* and *inv* mutant mice have shown that the incidence of *situs inversus* is ~50 and 100%, respectively. In contrast to wild-type mice, the expression of *nodal* in *inv/inv* mice is always in the right lateral plate mesoderm<sup>19,20</sup>. In *iv/iv* embryos, the

expression of *nodal* is occasionally in the left lateral plate mesoderm, but more often its pattern of expression is altered, being expressed for example only on the right, bilaterally, or not at all in the lateral plate mesoderm<sup>20</sup>. If *Pitx2* is an important factor in the left–right pathway and acts downstream of Nodal, its expression in these mouse mutants should parallel the observed patterns of *nodal* expression, which is what we found. In the embryos from an



**Figure 3** Ectopic expression of *Pitx2* affects left–right asymmetry of the heart and the gut. All embryos are shown from the ventral side. **a**, Wild-type stage 12 chick embryo with normal rightward heart looping (arrow). **b**, Stage 13 chick embryo infected with RCAS-*Pitx2* in the right lateral plate mesoderm at stage 4, showing a bilaterally symmetrical heart. Note the midline location of the bilaterally symmetrical heart with no bias towards the left or right side of the embryo. **c**, Stage 13 chick embryo that was infected with RCAS-*Pitx2* in right lateral plate mesoderm at stage 4 showing leftward heart looping (arrows). **d**, **e**, **f**, Higher magnification at stage 13 of *Pitx2* virus-infected hearts. **d**, Control embryo; heart looping is to the right. **e**, *Pitx2* infected embryo; bilaterally symmetrical heart without a defined looping towards either the left or right side of the embryo. **f**, *Pitx2* infected embryo; heart looping is to the left. **g**, Wild-type stage 45 *Xenopus* embryo with rightward looping heart and a counterclockwise coiled gut. **h**, Stage 45 *Xenopus* embryo that was injected with *Pitx2* at the 4-cell stage, with leftward heart looping and a clockwise coiled gut.

*inv* +  $\times$  *inv*/+ cross, *Pitx2* was expressed only in the right lateral plate mesoderm in 4/17 mice, which is the expected number of *inv*/*inv* embryos from such a cross (Fig. 4b). In *iv*/*iv* embryos, expression of *Pitx2* in the lateral plate mesoderm was random (Fig. 4c–f): in 20% (3/15) of them, *Pitx2* was expressed in the left lateral plate mesoderm, as in the wild type (Fig. 4a, c); in 20% (3/15), there was no *Pitx2* expression in the lateral plate mesoderm (Fig. 4d); in 20% (3/15), *Pitx2* expression was in the right lateral plate mesoderm (Fig. 4e); and in the remaining 40% (6/15), *Pitx2* was expressed bilaterally (Fig. 4f).

### Effect of dominant-negative ActRII on *Pitx2*

In the chick, activin- $\beta$ B affects left–right asymmetries<sup>8,9</sup>. The action of activin is effected by the asymmetric distribution of targets, presumably activin receptors. In mice, only deletion of activin receptor IIB results in defects in the left–right axis<sup>13,16–18</sup>. Injection of a dominant-negative form of the type II activin receptor in *Xenopus* altered cardiac looping and *Xnr-1* expression when injected on the left side of 16-cell embryos, presumably by interfering with normal Vg-1 signalling<sup>11</sup>.

To examine the relation between activin type II receptors, *Pitx2* expression, and visceral *situs*, we used a soluble dominant-negative receptor, ActRII-ECD, which contains the entire extracellular ligand-binding domain of the type II receptor and binds to activin with an affinity comparable to that of the intact receptor<sup>38</sup> (Fig. 5). When present in molar excess, this construct should therefore block activin signalling. Beads coated with ActRII-ECD (0.5 mg ml<sup>−1</sup>) were implanted into the left or right side of Hensen's node, or immediately above it at stage 4–5 in New culture<sup>10,39</sup>. When beads were implanted away from the node in either the right or left presumptive lateral plate mesoderm, about 30% of the embryos showed reversal of heart looping or bilateral hearts (Fig. 6g). The frequency of reversed heart looping and/or isomerism was higher when beads were placed in the midline (45%). As previously observed, control beads showed a 15% reversal of heart looping<sup>8</sup>. Implanting the dominant-negative activin receptor type II beads had no effect on *Shh* ( $n = 24$ ) expression, regardless of which side they were implanted on (data not shown), although they could induce *nodal* and *Pitx2* expression bilaterally in the lateral plate

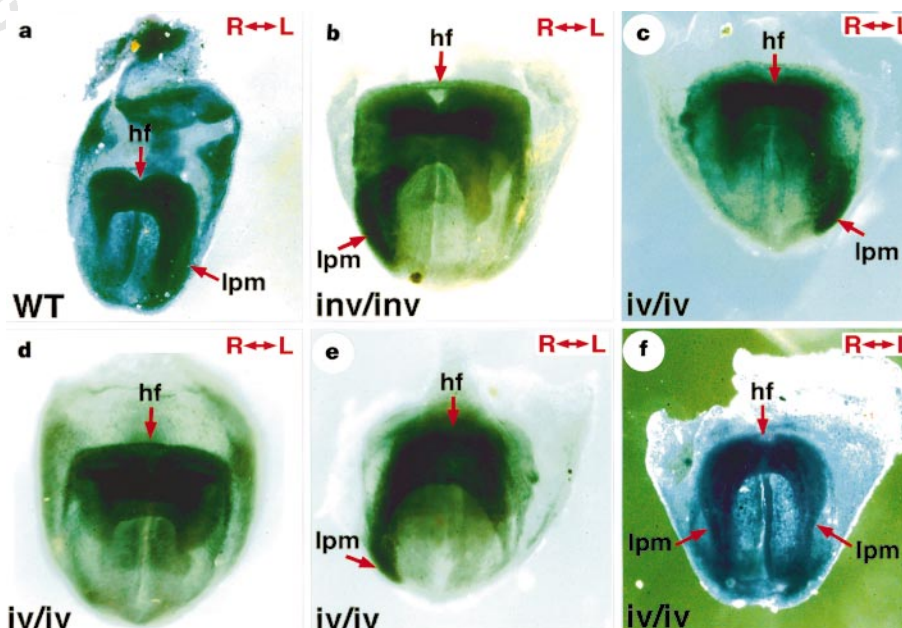
mesoderm (Fig. 6a–e). In a few embryos, *Pitx2* was not detected in either the left or the right lateral plate mesoderm, and in one embryo, *Pitx2* was expressed only in the right and not in the left lateral plate mesoderm (Fig. 6b). Taken together with the results of *nodal* misexpression and *Pitx2* expression in mutant mice, these results place *Pitx2* downstream of the left–right asymmetry pathway initiated by several members of the TGF- $\beta$  superfamily.

### Conservation of mechanism of L–R asymmetry

As the TGF- $\beta$  superfamily appears to be important in left–right patterning in several vertebrate species, we propose that these signalling molecules induce a specific transcription factor(s) that is critical for mediating left–right patterning information. We have presented evidence that *Pitx2* functions as the transcriptional mediator of left–right *situs*, being required and sufficient to induce the remainder of the downstream program needed to establish morphological asymmetries along this axis. Induction of *Pitx2* can be linked to signalling molecules critical for left–right patterning.

In the chick, asymmetric expression of *activin- $\beta$ B* on the right side of the node between stages 3 and 5<sup>+</sup> is the first molecular marker of left–right asymmetry<sup>9</sup>. The sequence of events probably involves *activin- $\beta$ B* functioning through *cAct-RIIB* to activate *cAct-RIIA* expression on the right side of Hensen's node, leading to down-regulation of *Shh* on the right side of the node. The asymmetric expression of *Shh* on the left side of the node at stage 4 leads to the induction, in the left lateral plate mesoderm, of *nodal*, the second member of the TGF- $\beta$  superfamily that regulates left–right asymmetry decisions in the chick<sup>8,9</sup>. *Shh* signalling is both necessary and sufficient for *nodal* induction in the chick<sup>26</sup>. In contrast to the chick model, targeted gene deletions of *Shh*, *activin- $\beta$ B* and the activin receptor IIA in the mouse have no apparent effect on visceral *situs*<sup>13,14,16,17</sup>. Although apparent differences between species may indicate that some components of the molecular machinery responsible for establishing left–right differences are not conserved<sup>26</sup>, we suggest that this reflects functional redundancies and that the basic flow of regulatory events is highly conserved among vertebrates.

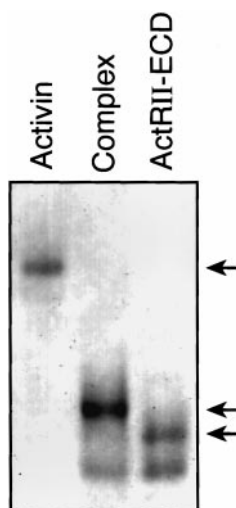
*Nodal* shows absolute conservation of expression in all vertebrate species that we have examined. Left-sided expression of *nodal* in the lateral plate mesoderm appears to be a prerequisite for establishing



**Figure 4** Expression of *Pitx2* is altered in *iv* and *inv* mice. Whole-mount *in situ* hybridization of 8.0 d.p.c. mouse embryos with *Pitx2*. **a**, *Pitx2* is expressed in the head-fold (hf) and left lateral plate mesoderm (lpm) in the wild-type mouse embryo. **b**, In *inv/inv* mice, *Pitx2* expression is observed in the head-fold and right

lateral plate mesoderm. **c–f**, Four different expression patterns of *Pitx2* in *iv/iv* mice: *Pitx2* in head-fold and left lateral plate mesoderm (**c**); *Pitx2* in head-fold but absent in lateral plate mesoderm (**d**); *Pitx2* in head-fold and in right lateral plate mesoderm (**e**); and, bilateral expression in the lateral plate mesoderm (**f**).





**Figure 5** ActRII-ECD binds activin. Non denaturing gel demonstrating the binding of activin and deglycosylated ActRII-ECD. Deglycosylation does not significantly affect the binding affinity. Lanes contain (from left): 1  $\mu$ g activin, 1  $\mu$ g activin mixed with 3  $\mu$ g ActRII-ECD, and 3  $\mu$ g ActRII-ECD; staining was with Coomassie blue. Arrows indicate protein bands.

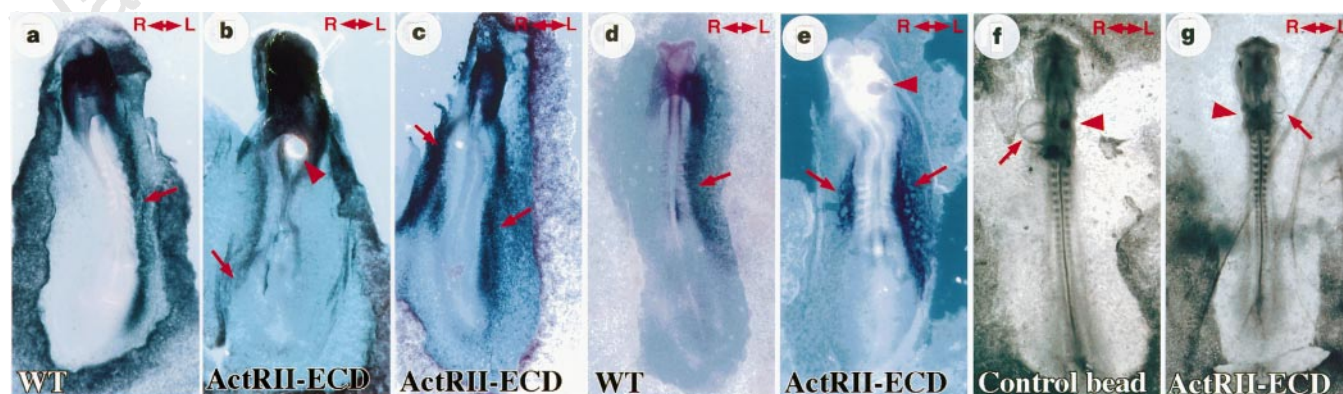
the normally invariant pattern of left-right asymmetries. Any variation from this normal *nodal* expression pattern, be it bilateral, absent, or only on the opposite (right) side, has a dramatic effect on the *situs* of the internal organs<sup>8,19,20,26,40</sup>. In mice, two closely related members of the TGF- $\beta$  family, *Lefty-1* and *Lefty-2*, are expressed in the left ventral midline of the gastrulating embryo and in the left lateral plate mesoderm, respectively<sup>41</sup>. *Lefty-1* and 2 expression in *inv*, *inv*, *Ft* and *no turning* mice parallel *nodal* expression<sup>22,23,41,42</sup>. Ablation of *Lefty-1* in the mouse causes bilateral expression of *nodal*<sup>43</sup>, suggesting that *Lefty-1* signalling participates in the mechanism that normally confines *nodal* expression to the left lateral plate mesoderm.

Our results in the chick using a dominant-negative activin type II receptor agree with the results of targeted gene deletion in the mouse<sup>13,16–18</sup> and from *Xenopus* microinjection<sup>11</sup>. The extracellular domain of the activin type II receptor used as a dominant-negative receptor in our experiments contains the entire ligand-binding domain of the type II activin receptor and can presumably interfere

with molecules that normally bind to IIA or IIB type receptors. An explanation of our data from the chick, in which there is bilateral *nodal* expression after applying this dominant-negative ActRII construct, is that *Lefty-1* signalling has been blocked in the chick, allowing the spread of *nodal* expression. The defects in heart *situs* in the chick could result from the bilateral expression of *nodal*; interference with Nodal function, or both. Taken together, our results indicate that Nodal, and perhaps *Lefty* or another similar factor, are key conserved signalling molecules that are required for establishing left–right asymmetry during embryogenesis. We suggest that, although there are apparent differences in the expression of some regulatory genes among model systems, the central mechanism for regulating left–right asymmetry has been widely conserved in vertebrate evolution.

We also suggest that *Pitx2* is the central induced effector of the programme mediating left–right patterning. In chick, *Xenopus* and mouse, *Pitx2* expression is first observed on the left side of the embryo in the lateral plate mesoderm. In the chick, *Pitx2* expression in the left lateral plate mesoderm at stage 8 occurs several hours after asymmetric *Shh* expression in the node at stage 4<sup>+</sup> and *nodal* expression in the left lateral plate mesoderm at stage 7. In mice with *situs inversus*, *Pitx2* expression parallels that of *nodal*, and in the chick and frog, misexpression of *nodal* (or *Shh*) induces ectopic *Pitx2* expression. Interference with signalling by TGF- $\beta$  family members using the dominant-negative activin type II receptor alters *Pitx2* expression and *situs*. In agreement with its predicted function, ectopic expression of *Pitx2* in the right side of the embryo affects left–right asymmetry of the heart and gut and reverses the direction of embryonic turning, resulting in phenotypes similar to those associated with *Shh* and *nodal* misexpression.

Although the different signalling molecules described here are transiently expressed and disappear before morphological asymmetries are visible, *Pitx2* expression is initiated at the onset of organogenesis and is maintained throughout embryogenesis. *Pitx2* is asymmetrically expressed both in the left lateral plate mesoderm, which appears to be the source of the inducing signal directing the morphological movements that give rise to left–right asymmetries, and in the lateral plate mesoderm derivatives that respond to these migration-inducing signals. The left–right instructive role of *Pitx2* is not restricted to the heart but is also required for correct *situs* of the gut and body rotation. Thus, to our knowledge, *Pitx2* is the first transcription factor demonstrated to regulate left–right asymmetry along the body axis, suggesting that *Pitx2* is an



**Figure 6** A dominant negative activin receptor alters expression of *nodal* and *Pitx2* and causes reversal of heart looping. All embryos are shown from the ventral side. **a**, Wild-type *Pitx2* expression. **b**, Ectopic *Pitx2* expression on the right lateral plate mesoderm (arrow) of an embryo with ActRII-ECD bead (arrowhead) implanted on the right side at stage 4. Note the absence of *Pitx2* expression on the left side. **c**, *Pitx2* expression in head-folds and bilaterally in lateral plate mesoderm (arrows) of an embryo with ActRII-ECD bead implanted on the right side at stage 4. **d**, Wild-type expression of *nodal* in the left lateral plate mesoderm (arrow). **e**,

Embryo in which an ActRII-ECD bead (arrow head) had been implanted on the right side at stage 4, showing ectopic *nodal* expression in the right lateral plate mesoderm (arrow). **f**, Normal rightward looping of the heart (arrow) in an embryo in which a control bead (arrowhead) had been implanted on the right side at stage 4. **g**, Leftward looping of heart (arrow) in an embryo in which an ActRII-ECD bead (arrowhead) had been implanted on the right side at stage 4. R  $\leftrightarrow$  L in the right top portion of the panels indicates the orientation of the embryo, R being right and L being left.



evolutionarily conserved downstream effector of the signalling cascade that establishes asymmetries along the entire left-right axis in vertebrates. □

# Methods

**cDNA cloning and *in situ* hybridization.** Stage 22–24 chick limb bud and stage 26 *Xenopus* cDNA libraries were screened at reduced stringency with full-length mouse P-OTX<sup>35</sup>. cDNAs were sequenced on both strands and the sequences deposited in Genbank (AF077092 and AF077767). Whole-mount *in situ* hybridization and sectioning of chick embryos (staged according to ref. 44) were done as described<sup>45</sup>. Antisense probes for *Xpitx2* and *cSnR* spanned the entire ORF, *cPitx2* the homeodomain and C terminus; *Nodal*<sup>8</sup>, *Shh*<sup>46</sup> and *Xnr-1* (ref. 40) were produced as described<sup>47</sup>.

**Retroviral infection.** RCASBP(A) retrovirus stocks were produced<sup>48</sup> containing full-length *cPitx2*, *Shh* or *cSnR* or mature chick *Nodal* fused with the BMP-4 proregion<sup>8</sup>. Embryos were infected by right-side blastoderm injection at stages 4–6 (ref. 49).

**Construction and microinjection of CDG-Xnr-1 and CDG-XPitx2.** The protein-coding regions of *Xnr-1* or *XPitx2* were PCR-amplified from a *Xenopus* gastrula library<sup>50</sup> (*Xnr-1*) or cDNA (*XPitx2*) and cloned into pCDG1 (ref. 51): 100 pg, plus 20 pg of CS2- $\beta$ -gal lineage tracer, were microinjected into one blastomere of 4-cell albino embryos. Fixation was at stage 24–26, and was followed by  $\beta$ -galactosidase staining<sup>47</sup>.

Surviving plasmid-injected embryos (63) were scored according to the location of the lineage tracer: 23 showed exclusive expression on the left and 21/23 (91%) had normal left *Pitx2* staining, whereas 2/23 (9%) had bilateral *Pitx2* expression. 21/63 showed exclusive expression of the lineage tracer on the right; 3/21 (14%) showed normal left expression of *Pitx2*, 12/21 (57%) showed bilateral *Pitx2* expression and 6/21 (29%) showed reversed right expression of *Pitx2*. 19/63 could not be scored owing to bilateral or absent lineage tracer expression. All 58 embryos injected with  $\beta$ -gal alone showed normal *Pitx2* expression.

Sense mRNA for microinjection was prepared as described<sup>47</sup>. *Pitx2* mRNA was injected into pigmented embryos for phenotypic or albino embryos for *in situ* analysis. For phenotypic analysis, embryos were fixed at stage 45, stained with the MF-20 antibody<sup>40</sup> and analysed by light or confocal laser scanning microscopy. *In situ* hybridization was done as described<sup>47</sup>.

**Protein purification and bead implantation.** Chick embryos were grown *in vitro* essentially as described<sup>39</sup> but with modifications<sup>10</sup>. The entire extracellular ligand-binding domain (ActRII-ECD; residues 1–116) of rat activin type II receptor was overexpressed in *Pichia pastoris*<sup>35</sup>. The protein was purified by nickel-affinity chromatography, followed by ion-exchange and size chromatography, and was shown to be monomeric in solution by analytical ultracentrifugation<sup>35</sup>. ActR-ECD was deglycosylated by endoglycosidase H (Glyko) and bound to activin (gift from W. W. Vale) with a dissociation constant of  $\sim 10$  nM. Affigel blue beads were soaked for 1 h with ActRII-ECD protein (0.5 mg ml<sup>-1</sup>) and implanted between the hypoblast and epiblast to the right, left or in the midline of stage 4 embryos.

Received 20 May; accepted 9 July 1998.

1. Burn, J. Disturbance of morphological laterality in humans. In *Biological Asymmetry and Handedness* (eds Bock, G. R. & Marsh, J.) 282–299 (Wiley, New York, 1991).
2. Kosaki, K. & Casey, B. Genetics of human left-right axis of malformations. *Sem. Cell Dev. Biol.* **9**, 89–99 (1998).
3. Supp, D. M., Bruickner, M. & Potter, S. S. Handed asymmetry in the mouse: Understanding how things go right (or left) by studying how they go wrong. *Sem. Cell Dev. Biol.* **9**, 77–87 (1998).
4. Splitt, M. P., Burn, J. & Goodship, J. Defects in the determination of left-right asymmetry. *J. Med. Genet.* **33**, 498–503 (1996).
5. Wood, W. B. Left-right asymmetry in animal development. *Annu. Rev. Cell Dev. Biol.* **13**, 53–82 (1997).
6. Lander, A., King, T. & Brown, N. A. Left-right development: Mammalian phenotypes and conceptual models. *Sem. Cell Dev. Biol.* **9**, 35–41 (1998).
7. Levin, M. & Mercola, M. The compulsion of chirality: towards understanding of left-right asymmetry. *Genes Dev.* **12**, 763–769 (1998).
8. Levin, M., Johnson, R. L., Stern, C. D., Kuehn, M. & Tabin, C. A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell* **82**, 803–814 (1995).
9. Levin, M. *et al.* Left/right patterning signals and the independent regulation different aspects of *situs* in the chick embryo. *Dev. Biol.* **189**, 57–67 (1997).
10. Stern, C. D. *et al.* Activin and its receptors during gastrulation and the later phases of mesoderm development in the chick embryo. *Dev. Biol.* **172**, 192–205 (1995).
11. Hyatt, B. A., Lohr, J. L. & Yost, H. J. Initiation of vertebrate left-right axis formation by maternal Vg1. *Nature* **384**, 62–65 (1996).
12. Hyatt, B. A. & Yost, H. J. The left-right coordinator: the role of Vg1 in organizing left-right axis formation. *Cell* **93**, 37–46 (1998).

13. Matzuk, M. M., Kumar, T. R. & Bradley, A. Different phenotypes for mice deficient in either activins or activin receptor type II. *Nature* **374**, 356–360 (1995).
14. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**, 407–413 (1996).
15. Matzuk, M. M. *et al.* Multiple defects and perinatal death in mice deficient in follistatin. *Nature* **374**, 360–363 (1995).
16. Matzuk, M. M. *et al.* Functional analysis of activins during mammalian development. *Nature* **374**, 354–356 (1995).
17. Vassalli, A., Matzuk, M. M., Gardner, H. A., Lee, K. F. & Jaenisch, R. Activin/inhibin  $\beta$ B subunit gene disruption leads to defects in eyelid development and female reproduction. *Genes Dev.* **8**, 414–427 (1994).
18. Oh, S. P. & Li, E. The signaling pathway mediated by the type IIB activin receptor controls axial patterning and lateral asymmetry in the mouse. *Genes Dev.* **11**, 1812–1826 (1997).
19. Collignon, J., Varlet, I. & Robertson, E. J. Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* **381**, 155–158 (1996).
20. Lowe, L. A. *et al.* Conserved left-right asymmetry of nodal expression and alterations in murine *situs inversus*. *Nature* **381**, 158–161 (1996).
21. Supp, D. M. *et al.* Mutation of an axonemal dynein affects left-right asymmetry in *inversus viscerum* mice. *Nature* **389**, 963–966 (1997).
22. Heymer, J., Kuehn, M. & Rüther, U. The expression pattern of nodal and Lefty in the mouse mutant Ft suggests a function in the establishment of handedness. *Mech. Dev.* **66**, 5–11 (1997).
23. Mello, P. G. *et al.* No turning, a mouse mutation causing left-right and axial patterning defects. *Dev. Biol.* **193**, 77–89 (1998).
24. Lohr, J. L., Danos, M. C. & Yost, H. J. Left-right asymmetry of a nodal-related gene is regulated by dorsoanterior midline structures during *Xenopus* development. *Development* **124**, 1467–1472 (1997).
25. Levin, M. Left-right asymmetry in vertebrate embryogenesis. *Bioessays* **19**, 287–296 (1997).
26. Pagán-Westphal, S. M. & Tabin, C. J. The transfer of left-right positional information during chick embryogenesis. *Cell* **93**, 25–35 (1998).
27. Isaac, A., Sargent, M. G. & Cooke, J. Control of vertebrate left-right asymmetry by a Snail-related zinc-finger gene. *Science* **275**, 1301–1304 (1997).
28. Srivastava, D., Cserjesi, P. & Olson, E. N. A subclass of bHLH proteins required for cardiac morphogenesis. *Science* **270**, 1995–1999 (1995).
29. Biben, C. & Harvey, R. P. Homeodomain factor Nkx2-5 controls left/right asymmetric expression of bHLH gene *eHAND* during murine heart development. *Genes Dev.* **11**, 1357–1369 (1997).
30. Semina, E. V. *et al.* Cloning and characterization of a novel bicoid-related homeobox transcription factor gene, RIEG, involved in Rieger syndrome. *Nature Genet.* **14**, 392–399 (1996).
31. Gage, P. J. & Camper, S. A. Pituitary homeobox 2, a novel member of the bicoid-related family of homeobox genes, is a potential regulator of anterior structure formation. *Hum. Mol. Genet.* **6**, 457–464 (1997).
32. Muccielli, M. L., Martinez, S., Pattyn, A., Goridis, C. & Brunet, J. F. Otx2, and Otx-related homeobox gene expressed in the pituitary gland and in a restricted pattern in the forebrain. *Mol. Cell Neurosci.* **8**, 258–271 (1996).
33. Hirofumi, A. *et al.* Identification and characterization of the ARP1 gene, a target for the human acute leukemia ALL1 gene. *Proc. Natl. Acad. Sci. USA* **95**, 4573–4578 (1998).
34. Lamonerie, T. *et al.* Ptx1, a bicoid-related homeobox transcription factor involved in transcription of the pro-opiomelanocortin gene. *Genes Dev.* **10**, 1284–1295 (1996).
35. Szeto, D. P., Ryan, A. K., O'Connell, S. M. & Rosenfeld, M. G. P-OTX: A PIT-1 interaction homeodomain factor expressed during anterior pituitary gland development. *Proc. Natl. Acad. Sci. USA* **93**, 7706–7710 (1996).
36. Lustig, K. D. *et al.* A *Xenopus* nodal-related gene that acts in synergy with noggin to induce complete secondary axis and notochord formation. *Development* **122**, 3275–3282 (1996).
37. Logan, M., Pagán-Westphal, S., Smith, D., Paganesi, L. & Tabin, C. J. The transcription factor of Ptx-2 mediates situs-specific morphogenesis in response to left-right asymmetric signaling. *Cell* (in press).
38. Donaldson, C. J., Vaughan, J. M., Corrigan, A. Z., Fischer, W. H. & Vale, W. W. Characterization of the soluble type II activin receptor extracellular domain. *Biochemistry* (submitted).
39. New, D. A. T. A new technique for the cultivation of the chick embryo *in vitro*. *J. Embryol. Exp. Morphol.* **3**, 326–331 (1955).
40. Sampath, K., Cheng, A. M. S., Frisch, A. & Wright, C. V. E. Functional differences among *Xenopus* nodal-related genes in left-right axis determination. *Development* **124**, 3293–3302 (1997).
41. Meno, C. *et al.* Two closely-related left-right asymmetrically expressed genes, *lefty-1* and *lefty-2*: their distinct expression domains, chromosomal linkage and direct neutralizing activity in *Xenopus* embryos. *Genes to Cells* **2**, 513–524 (1997).
42. Meno, C. *et al.* Left-right asymmetric expression of the TGF $\beta$ -family member *lefty* in mouse embryos. *Nature* **381**, 151–155 (1996).
43. Meno, C. *et al.* Left-1 is required for left-right determination as a regulator of Lefty-2 and Nodal. *Cell* (in press).
44. Hamburger, V. & Hamilton, H. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
45. Wilkinson, D. G. In *In Situ Hybridisation* (ed. Wilkinson, D. G.) (Oxford University Press, Oxford, 1993).
46. Yonei, S., Tamura, K., Ohsugi, K. & Ide, H. MRC-5 cells induce the AER prior to the duplicated pattern formation in chick limb bud. *Dev. Biol.* **170**, 542–552 (1995).
47. Blumberg, B. *et al.* An essential role for retinoid signaling in anteroposterior neural patterning. *Development* **124**, 373–379 (1997).
48. Vogel, A., Rodriguez, C. & Izpisua Belmonte, J. C. Involvement of FGF-8 in initiation, outgrowth and patterning of the vertebrate limb. *Development* **122**, 1737–1750 (1996).
49. Morgan, B. A., Izpisua Belmonte, J. C., Duboule, D. & Tabin, C. J. Targeted misexpression of Hox-4.6 in the avian limb bud causes apparent homeotic transformations. *Nature* **358**, 236–239 (1992).
50. Cho, K. W. Y., Blumberg, B., Steinbeisser, H. & De Robertis, E. M. The role of the *Xenopus* homeobox gene gooseoid. Molecular nature of Spemann's organizer. *Cell* **67**, 1111–1120 (1991).
51. Blumberg, B. *et al.* BXR, an embryonic orphan nuclear receptor activated by a novel class of endogenous benzoate metabolites. *Genes Dev.* **12**, 1269–1277 (1998).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank E. Leonardo for her technical skills and dedication; F. H. Gage for access to the confocal laser scanning microscope; D. Peterson for assistance with confocal imaging; C. Tabin for sharing unpublished results; C. V. E. Wright, C. Tabin and L. Erkmann and B. Eshelman for reagents and for discussion; J. Magallon and B. Eshelman for technical assistance; and L. Hooks and P. Myer for preparing the manuscript. S.Y.T. and K.T. were supported by the J.S.P.S. M.G.R. and R.M.E. are investigators of the Howard Hughes Medical Institute; J.C.I.B. is a Pew Scholar. J.G. is a HHMI predoctoral fellow. This work was supported by NIH grants to M.G.R., R.M.E. and J.C.I.B., by a G. Harold and Leila Y. Mathers Charitable Foundation grant to R.M.E., S.C., and J.C.I.B.

Correspondence and requests for materials should be addressed to J.C.I.B. (e-mail: [belmonte@salk.edu](mailto:belmonte@salk.edu)).

# Emergence of magnetic flux on the Sun as the cause of a 158-day periodicity in sunspot areas

R. Oliver\*, J. L. Ballester\* & F. Baudin†

\* Departament de Física, Universitat de les Illes Balears, E-07071 Palma de Mallorca, Spain

† Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

The temporal behaviour of solar activity (as manifested in sunspots) has long been debated. The 11-year periodicity in the total number of sunspots is well established observationally, as is a periodicity of 152–158 days in the occurrence of high-energy solar flares that was seen during cycle 21 (refs 1–7). The cause of the latter periodicity is not clear, although several mechanisms have been proposed<sup>8–10</sup>. Here we report a time–frequency analysis, using the wavelet technique, of sunspot areas between 1874 and 1993, which reveals a 158-day periodicity coincident with that of energetic solar flares. The signature of this periodicity is strongest in cycle 19, which was the most intense cycle of the century. The periodicity disappears after cycle 21. The analysis shows that the 158-day periodicity in both high-energy solar flares and sunspots is related to a periodic emergence of magnetic flux which only appears near the maxima of some solar cycles.

Most of the mechanisms proposed to explain solar flares, especially the most energetic flares, accept as a prerequisite the emergence of magnetic flux<sup>11,12</sup> which triggers the destabilization of active regions. So we have put forward the hypothesis<sup>13</sup> that the periodic increase in the occurrence rate of energetic flares is related to a periodic emergence of magnetic flux through the photosphere, giving rise to a periodic variation of the total sunspot area on the Sun's surface.

Such flux emergence and the subsequent destabilization has been detected, for instance, within the active region underlying the X-class flare on 15 November 1991<sup>14</sup>. However, an observational confirmation of the above hypothesis is almost impossible because of the need for a continuous and multiwavelength (using ground-based and spatial observatories) coverage of the Sun's atmosphere. On the other hand, a study<sup>15</sup> of the magnetic flux variations during solar cycle 21 revealed the existence of quasi-periodic pulses or episodes of enhanced magnetic activity. The duration of the pulses is ~5 rotations during the years around maximum activity, the epoch in which the flare periodicity appears, and comparison with magnetic field maps indicates that those pulses of activity correspond to the occurrence of complex active regions containing large sunspots<sup>16</sup>. However, such synoptic studies are unable to prove the existence of a coincidence in time and in frequency between the emergence of magnetic flux and the occurrence of the periodicity in high-energy solar flares.

Also, we note that the periodicity around 152–158 days has been detected in other indicators of solar activity<sup>13,17–19</sup>, which suggests that it is associated preferentially with photospheric regions of compact magnetic-field structures related to sunspots rather than to plages.

To confirm the above hypothesis we need to prove that there has been a time and frequency coincidence between the occurrence of the periodicity in high-energy flares and in sunspot areas. To this end, we have examined records of sunspot areas, looking for the presence of a periodicity around 152–158 days, with an interval of occurrence coincident with that of high-energy solar flares (around the maximum of cycle 21).

Wavelet analysis is perfectly suited for this job. Classical Fourier analysis allows the study of a signal only in the frequency domain,

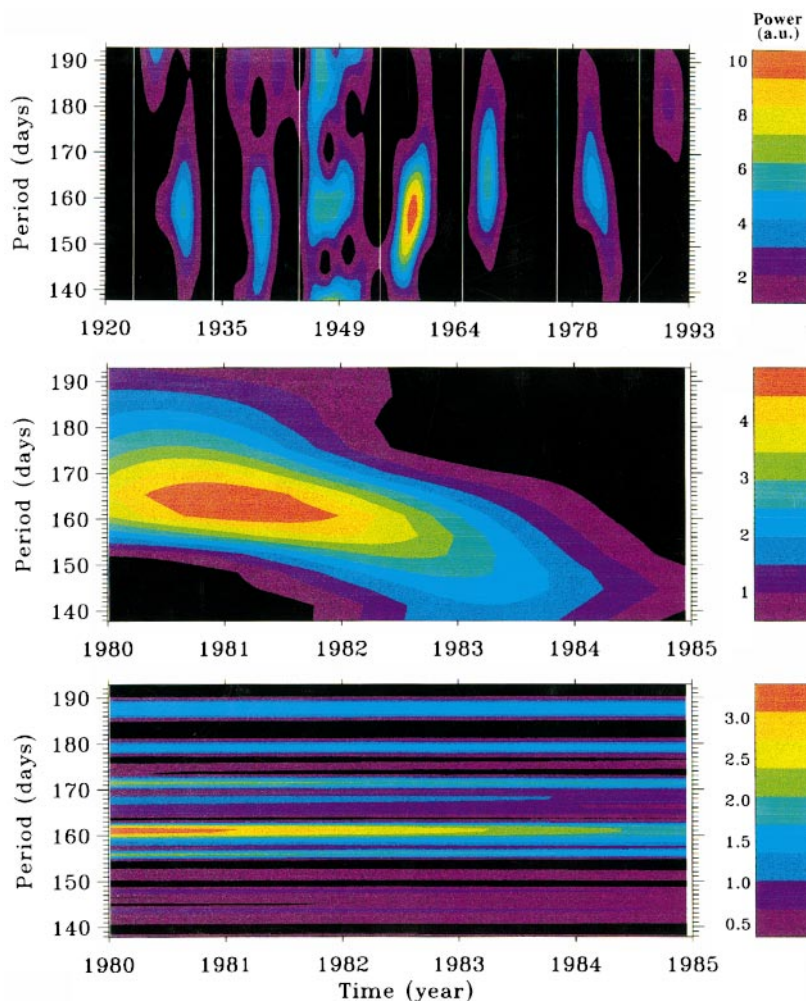
whereas wavelet analysis yields information in both time and frequency domains. That is why this technique has been used in many fields of physics in order to study the temporal behaviour of oscillatory signals. We have applied wavelet analysis to a time series consisting of daily sunspot areas between 1874 and 1993 in order to study the temporal variation with timescales of ~160 days. The data have been taken from the Greenwich Photoheliographic Results (1874–1982) and from the United States Air Force (1983 to July 1993); see Fig. 1 legend for details.

We have determined the epochs of appearance of the periodicity, and have found two new features (Fig. 1, top); first, that the periodicity has been present only around the maximum of solar activity in cycles 16 to 21, and second, that the power of the periodicity started growing at cycle 16, peaked at cycle 19 and decreased in subsequent cycles, completely disappearing after cycle 21. These features show that the periodicity does not operate during the whole solar cycle, and that the magnetic-flux emergence presents periodic episodes—around the maximum of solar activity of some solar cycles—whose strength seems to be correlated with the solar-cycle strength, because solar cycle 19 has been the most intense of this century.

The middle panel of Fig. 1 focuses on sunspot signal during the period when energetic solar flares were first studied<sup>1</sup> (1980–84), and shows that a similar periodicity was present in sunspot areas during this epoch. According to this figure, power around 161 days substantially decreases after 1984, in good agreement with the behaviour displayed by the record of energetic flares detected with HXRBS<sup>2</sup>, which shows a periodicity around 158 days between 1980 and 1984. So our analysis reveals the existence of a periodic variation in the emergence of magnetic flux which is in temporal coincidence with the periodicity found in high-energy solar flares, which suggests a close relationship between them. Further evidence for such relationship comes from two additional facts. First, power around 161 days was completely absent during the maximum of solar activity in cycle 22 (Fig. 1, top), in coincidence with the non-detection of the periodicity in energetic solar flares<sup>20</sup> and sunspots<sup>21</sup> during this epoch. Second, the power spectrum of proton flares<sup>4</sup> in the temporal interval 1958–71, which comprises the maxima of solar cycles 19 and 20, shows that the periodicity has been active during this interval.

We then used a higher frequency resolution in order to establish more precisely the value of the period in the temporal interval 1980–1985. (Fig. 1, bottom). The effect of the higher frequency resolution is to confine the power to a narrow band around 161 days. The period found for sunspot areas is in striking coincidence with that obtained for energetic solar flares.

Wavelet analysis of sunspot areas shows that the periodicity detected in high-energy solar flares is due to a periodic emergence of magnetic flux which is coincident in both time and frequency. This finding also gives support to the mechanisms which suggest that the factor triggering the destabilization of active regions, and which leads to the occurrence of energetic solar flares, is the emergence of new magnetic flux. Most of the theoretical models trying to explain the spatial and temporal behaviour of solar activity accept that magnetic flux is brought to the Sun's surface in the form of magnetic flux tubes which rise through the solar convection zone. The new features that we have found in the temporal behaviour of the periodicity add a further challenge to such models, as they need to be able to explain why the periodicity only appears around the maximum of the solar cycle, why its strength is correlated with the strength of the solar cycle, and why it disappears after some solar cycles. High-energy solar flares are also of great interest because they can produce severe effects on the Earth, such as power blackouts, disruption of communications, and damage to satellites and manned space flights. Understanding the cause of the episodes of periodic emergence of magnetic flux could help us to forecast increases in the occurrence of energetic solar flares, and so help to prevent the effects of such flares on the Earth and its environment. □



**Figure 1** Time/period diagrams from the wavelet analysis (power in arbitrary units) of daily sunspot areas between 1874 and 1993. As opposed to Fourier transform, which can be described as a decomposition of the signal into sine waves of constant amplitude and infinite duration in time, wavelet analysis uses functions which are band-limited in both time and frequency. Because we are dealing with a simple oscillatory signal, we have used the Morlet wavelet<sup>22</sup>, which is a sine wave windowed in time by a gaussian function. The resolution (the width of the window) in time and frequency has to be carefully chosen, depending on the aim: good temporal resolution is necessary to localize the power maxima in time, whereas good frequency resolution is required to determine the frequencies to which these maxima correspond. There is of course a trade-off between the resolution in each domain, defined by the uncertainty relation<sup>23</sup>. The use of the Morlet wavelet allows the best trade-off between time and frequency resolution,

as the gaussian function is its own Fourier transform. Top, analysis of the whole time series with a frequency resolution of 4 nHz, showing the temporal behaviour of the periodicity in the signal. Vertical white lines mark the epochs of minimum solar activity. Large values of power around 160 days can be seen in cycles 16 to 21, peaking at cycle 19. Middle, expanded-scale view of the period 1980–85, allowing the localization in time of the power maximum during cycle 21. Bottom, analysis of the same period as the middle panel, but with a frequency resolution of 0.5 nHz, to determine the frequency corresponding to the power maximum. The small amount of power present at other periods corresponds to peaks in the Fourier power spectrum which are well below the confidence level. Sunspot data were provided by the National Geophysical Data Center of the National Oceanic and Atmospheric Administration, and may be downloaded at <http://www.ngdc.noaa.gov/stp/stp.html>

Received 9 March; accepted 12 May 1998.

1. Rieger, E. *et al.* A 154-day periodicity in the occurrence of hard solar flares? *Nature* **312**, 623–625 (1984).
2. Dennis, B. R. Solar hard X-ray bursts. *Sol. Phys.* **100**, 465–490 (1985).
3. Bogart, R. S. & Bai, T. Confirmation of a 152 day periodicity in the occurrence of solar flares inferred from microwave data. *Astrophys. J.* **299**, L51–L55 (1985).
4. Bai, T. & Cliver, E. W. A 154 day periodicity in the occurrence rate of proton flares. *Astrophys. J.* **363**, 299–309 (1990).
5. Dröge, W. *et al.* A 153 day periodicity in the occurrence of solar flares producing energetic interplanetary electrons. *Astrophys. J. Suppl. Ser.* **73**, 279–283 (1990).
6. Kile, J. N. & Cliver, E. W. A search for the 154 day periodicity in the occurrence rate of solar flares using Ottawa 2.8 GHz burst data, 1955–1990. *Astrophys. J.* **370**, 442–448 (1991).
7. Verma, V. K., Joshi, G. C. & Paliwal, D. C. Study of periodicities of solar nuclear gamma ray flares and sunspots. *Sol. Phys.* **138**, 205–208 (1992).
8. Bai, T. & Sturrock, P. A. The 154-day and related periodicities of solar activity as subharmonics of a fundamental period. *Nature* **350**, 141–142 (1991).
9. Sturrock, P. A. & Bai, T. Search for evidence of a clock related to the solar 154 day complex of periodicities. *Astrophys. J.* **397**, 337–346 (1992).
10. Bai, T. & Sturrock, P. A. Evidence for a fundamental period of the Sun and its relation to the 154 day complex of periodicities. *Astrophys. J.* **409**, 476–486 (1993).
11. Forbes, T. G. Magnetic reconnection in solar flares. *Geophys. Astrophys. Fluid Dyn.* **62**, 15–36 (1991).
12. Priest, E. R. in *Basic Plasma Processes on the Sun* (eds Priest, E. R. & Krishnan, V.) 271–291 (Kluwer Academic, Dordrecht, 1990).

13. Carbonell, M. & Ballester, J. L. A short-term periodicity near 155 day in sunspot areas. *Astron. Astrophys.* **238**, 377–381 (1990).
14. Canfield, R. C. *et al.* in *X-Ray Solar Physics From Yohkoh* (eds Uchida, Y., Watanabe, T., Shibata, K. & Hudson, H. S.) 153–156 (Universal Academy, Tokyo, 1993).
15. Rabin, D. M., DeVore, C. R., Sheeley, N. R., Harvey, K. L. & Hoeksema, J. T. in *Solar Interior and Atmosphere* (eds Cox, A. C., Livingston, W. C. & Matthews, M. S.) 781–843 (Univ. Arizona Press, Tucson, 1991).
16. Bai, T. Distribution of flares on the Sun: Superactive regions and active zones of 1980–1985. *Astrophys. J.* **314**, 795–807 (1987).
17. Lean, J. L. & Brueckner, G. E. Intermediate-term solar periodicities: 100–500 days. *Astrophys. J.* **337**, 568–578 (1989).
18. Lean, J. L. Evolution of the 155 day periodicity in sunspot areas during solar cycles 12 to 21. *Astrophys. J.* **363**, 718–727 (1990).
19. Carbonell, M. & Ballester, J. L. The periodic behaviour of solar activity: the near 155 day periodicity in sunspot areas. *Astron. Astrophys.* **255**, 350–362 (1992).
20. Bai, T. The 77 day periodicity in the flare rate of cycle 22. *Astrophys. J.* **388**, L69–L72 (1992).
21. Oliver, R. & Ballester, J. L. Short-term periodicities in sunspot areas during solar cycle 22. *Sol. Phys.* **156**, 145–155 (1990).
22. Grossmann, A. & Morlet, J. Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* **15**, 723–736 (1984).
23. Bracewell, R. N. *The Fourier Transform and its Applications* 160 (McGraw-Hill, New York, 1978).

Correspondence and requests for materials should be addressed to J.L.B. (e-mail: [dfslib0@ps.uib.es](mailto:dfslib0@ps.uib.es)).



# Production of O<sub>2</sub> on icy satellites by electronic excitation of low-temperature water ice

M. T. Sieger, W. C. Simpson & T. M. Orlando

W. R. Wiley Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, MS K8-88, PO Box 999, Richland, Washington 99352, USA

The signature of condensed molecular oxygen has been reported in recent optical-reflectance measurements of the jovian moon Ganymede<sup>1</sup>, and a tenuous oxygen atmosphere has been observed on Europa<sup>2</sup>. The surfaces of these moons contain large amounts of water ice, and it is thought that O<sub>2</sub> is formed by the sputtering of ice by energetic particles from the jovian magnetosphere<sup>3–8</sup>. Understanding how O<sub>2</sub> might be formed from low-temperature ice is crucial for theoretical and experimental simulations of the surfaces and atmospheres of icy bodies in the Solar System. Here we report laboratory measurements of the threshold energy, cross-section and temperature dependence of O<sub>2</sub> production by electronic excitation of ice in vacuum, following electron-beam irradiation. Molecular oxygen is formed by direct excitation and dissociation of a stable precursor molecule, rather than (as has been previously thought) by diffusion and chemical recombination of precursor fragments. The large cross-section for O<sub>2</sub> production suggests that electronic excitation plays an important part in the formation of O<sub>2</sub> on Ganymede and Europa.

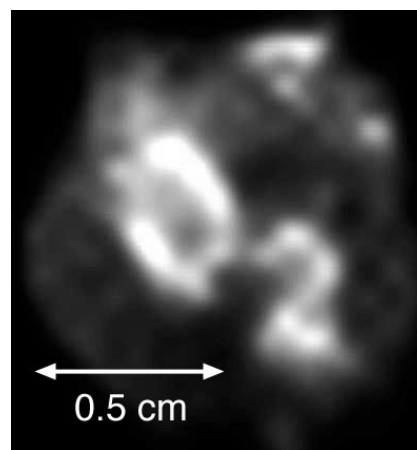
Janzerotti *et al.*<sup>9</sup> were the first to use data from space-probe measurements of the jovian magnetosphere to examine energetic ion sputtering of the galilean satellites; their work was subsequently extended by Johnson *et al.*<sup>10,11</sup>. More recently, Ip *et al.*<sup>8</sup> used data from the Galileo spacecraft to assess ion sputtering of Ganymede. The reported heavy-ion and proton bombardment flux is between 10<sup>8</sup> and 10<sup>11</sup> particles per cm<sup>2</sup> per s; the vacuum ultraviolet photon and electron flux ranges are expected to be similar. All of these studies concluded that the sputtering of water ice plays a central role in the evolution of the icy moons, their atmospheres, and the plasma composition of the jovian magnetosphere.

It is well known that the sputtering of low-temperature ice by fast, light ions is determined by the electronic excitations produced in the ice, rather than direct momentum transfer via collisions of the ions with water molecules<sup>5,12</sup>. This phenomenon, termed 'electronic sputtering', includes all loss of material from the ice as a result of electronic excitation by photons, electrons or fast ions. Fast-ion bombardment of ice is important on Ganymede and Europa, and most of the ion energy is deposited into direct excitations and secondary electron production<sup>8–12</sup>. Secondary electrons can then cause further excitations, forming an 'electronic cascade' by which a single incident ion produces thousands of electronic excitations. Excited water molecules may then break apart<sup>13,14</sup>, ejecting molecular fragments from the surface or trapping them in the ice. To assess the role of electronic excitations in the production of molecular oxygen, we have studied low-energy (<100 eV) electron bombardment of thin ice films. Thin water-ice films are relevant to conditions present on icy outer Solar System bodies, because the surfaces of these bodies are ice-enriched. The production and redistribution of gas-phase water molecules by meteorite impacts, sputtering and thermal desorption would suggest that even exposed rocks are coated with thin layers of water ice or frost. The use of electron bombardment as an excitation source has many advantages over high-energy ions, because the low excitation density, absence of direct momentum-transfer processes, and negligible local heating

allow us to examine the purely electronic processes involved in O<sub>2</sub> formation. Once the mechanism of O<sub>2</sub> formation in ice is understood, it can then be applied to a diverse range of excitation sources (for example, photons, electrons and ions) and to diverse environments, such as the jovian moons, comets, icy grains, and other outer Solar System bodies.

The experiments were performed in an ultra-high vacuum system (~10<sup>–10</sup> torr) equipped with a pulsed low-energy electron gun and a quadrupole mass spectrometer. The 5–100 eV mono-energetic electron beam supplied a time-averaged electron flux of ~6 × 10<sup>13</sup> electrons per cm<sup>2</sup> per s. Amorphous ice samples were prepared by vapour-depositing thin (~150 Å) films of D<sub>2</sub>O on a clean Pt(111) substrate at 110 K, at a rate of 4–8 monolayers per min. D<sub>2</sub>O was used instead of H<sub>2</sub>O to help distinguish ejected D<sub>2</sub>O from background gases. Other than an isotope shift in the temperature scale, it has been shown in sputtering experiments<sup>15</sup> that D<sub>2</sub>O and H<sub>2</sub>O ice behave similarly. The desorption products were detected with the quadrupole mass spectrometer in an 'electron beam on–beam off' mode which allowed for background subtraction. Other details of the apparatus and sample preparation have been published elsewhere<sup>14</sup>.

Large amounts of O<sub>2</sub> (typically one molecule per 1,000 incident electrons) were generated from our ice samples during electron bombardment. The electron energy threshold was 10 ± 2 eV, corresponding to valence electronic excitation of ice, and the yield generally increased with ice temperature. Oxygen is not produced immediately on bombardment; an 'incubation dose' is required before the yield becomes appreciable. A similar incubation dose has been observed in ion- and vacuum ultraviolet photon-induced sputtering of ice<sup>16,17</sup>, and implies a two-step (precursor-mediated) sputtering mechanism. To demonstrate that a precursor is needed to form O<sub>2</sub>, and to investigate its lifetime, we exposed regions of a fresh ice sample, selected to write out "O<sub>2</sub>" using the electron beam steering optics, to a ~10<sup>15</sup> cm<sup>–2</sup> dose of 50 eV (significantly above threshold) electrons. The O<sub>2</sub> yield was then measured as the electron beam was rastered quickly over the entire sample, forming an image of the O<sub>2</sub>-producing regions (Fig. 1). The contrast between the irradiated and non-irradiated regions shows that O<sub>2</sub> is produced via

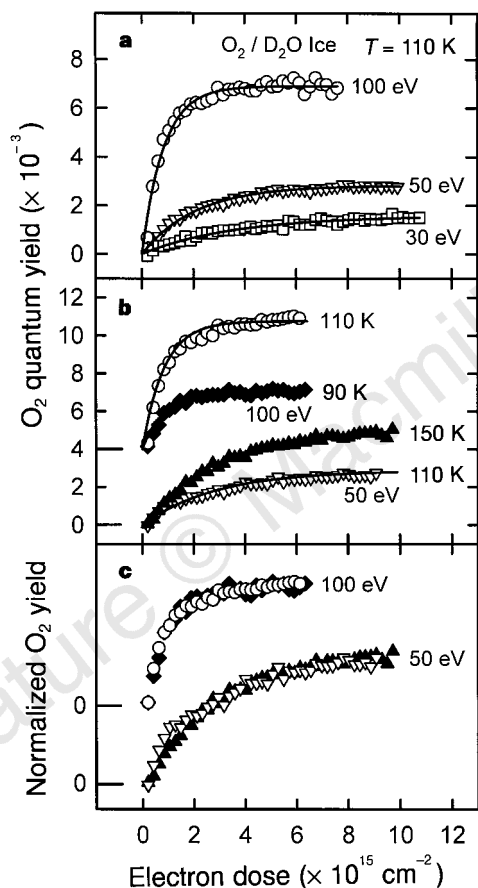


**Figure 1** Image of the O<sub>2</sub> yield as a function of electron-beam spot position on an ice thin-film sample. Regions selected to spell out "O<sub>2</sub>" were exposed to a ~10<sup>15</sup> cm<sup>–2</sup> dose of 50-eV electrons, at a sample temperature of 120 K. The sample is 1 cm in diameter, and the electron beam has a spot size of ~0.4 mm<sup>2</sup>. After the initial exposure, the electron beam was rastered over the sample, and the O<sub>2</sub> yield was measured as a function of the beam spot position. Bright and dark areas correspond to high and low O<sub>2</sub> yields, respectively. The circular profile of the sample is evident, as are the symbols marking the pre-bombarded areas. The contrast between the irradiated and non-irradiated regions demonstrates that O<sub>2</sub> is produced by a stable precursor. The elapsed time between 'writing' and 'reading' was ~1 h; the precursor species must therefore be stable on at least this timescale.

a precursor which is stable on the timescale of at least an hour at 120 K, suggestive of a stable molecular species, instead of a short-lived reactive species (such as atomic O).

By observing how the incubation dose depends on electron energy and ice temperature, we can extract detailed information about the O<sub>2</sub> formation mechanism. Measurements of the O<sub>2</sub> yield as a function of electron dose are shown in Fig. 2 for selected electron energies and ice temperatures. The yield is initially zero, rises with increasing dose, and eventually reaches a steady-state value. There are only a limited number of kinetic models which are consistent with our data. In all cases, the steady-state yield is limited by the precursor formation rate, whereas the dose needed to reach steady state is determined by the rate at which precursors form O<sub>2</sub>.

The data in Fig. 2 show that the steady-state O<sub>2</sub> yield increases with both electron energy and ice temperature. The dose needed to reach steady state depends on electron energy (Fig. 2a), but is independent of temperature (Fig. 2b, c). The data from Fig. 2b have



**Figure 2** O<sub>2</sub> yield versus electron dose. **a**, O<sub>2</sub> quantum yield (molecules per incident electron) as a function of total electron dose for selected electron energies, at an ice temperature of 110 K. Open circles, 100 eV; open triangles, 50 eV; open squares, 30 eV. The O<sub>2</sub> yield increases with electron energy, and the dose needed to reach steady state decreases with increasing energy. **b**, O<sub>2</sub> quantum yield as function of total electron dose, for selected ice temperatures and electron energies (the 100 eV data has been offset by  $+4 \times 10^{-3}$  for clarity). Filled diamonds, 100 eV at 90 K; filled triangles, 50 eV at 150 K; other symbols as in **a**. The yield increases with temperature, but the dose needed to reach steady state does not. **c**, The data from normalized at their steady-state values and plotted together, to show more clearly the independence of the rise to steady state on temperature. The 100-eV data are offset for clarity. All data were acquired at a fixed electron flux ( $6 \times 10^{13} \text{ cm}^{-2} \text{ s}^{-1}$ ) and measured as a function of time. Once the yield has saturated, it is linear in the instantaneous electron flux, and the energy threshold is  $10 \pm 2 \text{ eV}$  (referenced to the vacuum level). Plotted with the data are best-fit curves generated with the precursor dissociation model (solid lines).

been normalized at their steady-state values and plotted together in Fig. 2c to show more clearly the independence of the incubation dose on temperature. From these observations we can draw two conclusions: the precursor formation mechanism depends on temperature, while the process by which precursors are converted to O<sub>2</sub> does not depend on temperature in the temperature range studied.

The increase in the sputter yield with dose and temperature has traditionally been interpreted as arising from the build-up and thermal diffusion of radical precursors which chemically react to form O<sub>2</sub>. This hypothesis, however, cannot explain our results, as the process by which precursors form O<sub>2</sub> is temperature independent. The observed incubation-dose behaviour is also inconsistent with thermal diffusion in the precursor formation step, as diffusion models predict a quadratic rise in the yield near zero dose, while the data is clearly linear.

A two-step 'precursor dissociation' mechanism reproduces all the salient features of the data, and suggests to us a new interpretation of thermal effects in electronic sputtering of ice. A water molecule is first excited by an electron, photon or ion impact. Small changes in the structure or dynamic relaxation of ice can cause the excited state lifetime, and therefore the dissociation probability, to increase with temperature. We propose that the temperature dependence of O<sub>2</sub> production is in the lifetime of the electronic excited states leading to dissociation of water molecules. There is recent evidence that excited-state lifetimes in ice do increase with temperature<sup>14,18</sup>, in a manner suggestive of changes in the local hydrogen bonding. The dissociation fragments (O, OH) then reactively scatter to form a stable precursor molecule, possibly HO<sub>2</sub> or H<sub>2</sub>O<sub>2</sub>, which are known to be produced in irradiated ice<sup>6,19,20</sup>. A second electronic excitation then directly dissociates the precursor molecule to form O<sub>2</sub>.

Estimates of the precursor creation and dissociation cross-sections obtained by fitting the data to our model (Fig. 2, solid lines) are listed in Table 1, along with the measured cross-section for the stimulated desorption of D<sub>2</sub>O. We estimate that the steady-state O<sub>2</sub> production rate at Ganymede from secondary electrons alone should exceed  $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ . The total (secondaries plus direct excitation) production rate is likely to be at least an order of magnitude higher. The measured threshold of  $\sim 10 \text{ eV}$  is accessible to Lyman- $\alpha$  photons (10.2 eV), and so O<sub>2</sub> may also be produced by direct solar illumination. The measured cross-sections imply that the concentration of precursors in the near-surface region of ice is  $< 1\%$ , and so may elude all but the most sensitive spectroscopic detection<sup>21</sup>. However, O<sub>2</sub> may accumulate in the ice or become trapped at mineral/ice interfaces, giving rise to the observed absorption spectra<sup>1,7</sup>. Subsequent reactive interactions of water fragments with trapped O<sub>2</sub> can then form ozone (O<sub>3</sub>), which has been observed on Ganymede<sup>22,23</sup>.

The source of the temperature dependence implies that changes in the local electronic structure are more important than thermal diffusion for chemical processes in the electronic sputtering of ice. A prediction of our model is that the precursor concentration at steady state is a function only of temperature, and not electron/ion/photon flux, suggesting that the precursor concentration should be directly proportional to surface temperature; this is in agreement with recent observations showing that on Ganymede, more con-

**Table 1** O<sub>2</sub> precursor creation and dissociation

| Desorption product                    | Cross-section (cm <sup>2</sup> ) |                       |                       |
|---------------------------------------|----------------------------------|-----------------------|-----------------------|
|                                       | $E_i = 100 \text{ eV}$           | $E_i = 50 \text{ eV}$ | $E_i = 30 \text{ eV}$ |
| D <sub>2</sub> O                      | $1 \times 10^{-18}$              |                       |                       |
| O <sub>2</sub> precursor creation     | $2 \times 10^{-18}$              | $8 \times 10^{-19}$   | $5 \times 10^{-19}$   |
| O <sub>2</sub> precursor dissociation | $3 \times 10^{-16}$              | $1 \times 10^{-16}$   | $8 \times 10^{-17}$   |

Shown are estimated cross-sections (to within an order of magnitude) for D<sub>2</sub>O-stimulated desorption, and O<sub>2</sub> precursor formation and dissociation for selected electron energies  $E_i$ , at a sample temperature of 110 K. The O<sub>2</sub> precursor creation and precursor dissociation cross-sections are obtained from a fit of the precursor dissociation model to the data of Fig. 2. The total O<sub>2</sub> production cross-section at steady state is essentially the O<sub>2</sub> precursor creation cross-section, as the rate-limiting step in O<sub>2</sub> formation is the precursor formation rate.

densed O<sub>2</sub> is present at the equatorial latitudes than at the colder poles<sup>4</sup>. The mechanism proposed here is applicable to electronic excitations caused by essentially any radiation source (electrons, ions or photons), provides a quantitative understanding of O<sub>2</sub> formation in ice, and indicates that oxygen is probably also produced on other icy outer Solar System bodies. □

Received 8 December 1997; accepted 27 April 1998.

1. Spencer, J. R., Calvin, W. M. & Person, M. J. Charge-coupled device spectra of the Galilean satellites: Molecular oxygen on Ganymede. *J. Geophys. Res.* **100**, 19049–19056 (1995).
2. Hall, D. T., Strobel, D. F., Feldman, P. D., McGrath, M. A. & Weaver, H. A. Detection of an oxygen atmosphere on Jupiter's moon Europa. *Nature* **373**, 677–679 (1995).
3. Calvin, W. M., Johnson, R. E. & Spencer, J. R. O<sub>2</sub> on Ganymede: Spectral characteristics and plasma formation mechanisms. *Geophys. Res. Lett.* **23**, 673–676 (1996).
4. Calvin, W. M. & Spencer, J. R. Latitudinal distribution of O<sub>2</sub> on Ganymede: Observations with the Hubble Space Telescope. *Icarus* **130**, 505–516 (1997).
5. Johnson, R. E. Sputtering of ices in the outer solar system. *Rev. Mod. Phys.* **68**, 305–312 (1996).
6. Johnson, R. E. & Quickenden, T. I. Photolysis and radiolysis of water ice on outer solar system bodies. *J. Geophys. Res.* **102**, 10985–10996 (1997).
7. Vidal, R. A., Bahr, D., Baragiola, R. A. & Peters, M. Oxygen on Ganymede: Laboratory studies. *Science* **276**, 1839–1842 (1997).
8. Ip, W.-H., Williams, D. J., McEntire, R. W. & Mauk, B. Energetic ion sputtering effects at Ganymede. *Geophys. Res. Lett.* **24**, 2631–2634 (1997).
9. Lanzerotti, L. J., Brown, W. L., Poate, J. M. & Augustyniak, W. M. On the contribution of water products from Galilean satellites to the Jovian magnetosphere. *Geophys. Res. Lett.* **5**, 155–158 (1978).
10. Johnson, R. E. *Energetic Charged-particle Interactions with Atmospheres and Surfaces* (Springer, Berlin, 1990).
11. Johnson, R. E., Lanzerotti, L. J., Brown, W. L. & Armstrong, T. P. Erosion of Galilean satellite surfaces by Jovian magnetosphere particles. *Science* **212**, 1027–1030 (1980).
12. Brown, W. L., Lanzerotti, L. J., Poate, J. M. & Augustyniak, W. M. "Sputtering" of ice by MeV light ions. *Phys. Rev. Lett.* **40**, 1027–1030 (1978).
13. Kimmel, G. A. & Orlando, T. M. Low-energy (5–120 eV) electron-stimulated dissociation of amorphous D<sub>2</sub>O ice: D<sup>(2)S</sup>, O<sup>(2)P<sub>2,1,0</sub></sup> and O<sup>(2)D<sub>2</sub></sup> yields and velocity distributions. *Phys. Rev. Lett.* **75**, 2606–2609 (1995).
14. Sieger, M. T., Simpson, W. C. & Orlando, T. M. Electron-stimulated desorption of D<sup>+</sup> from D<sub>2</sub>O ice: Surface structure and electronic excitations. *Phys. Rev. B* **56**, 4925–4937 (1997).
15. Brown *et al.* Erosion and molecule formation in condensed gas films by electronic energy loss of fast ions. *Nucl. Inst. Methods* **198**, 1–8 (1982).
16. Reimann, C. T. *et al.* Ion-induced molecular ejection from D<sub>2</sub>O ice. *Surf. Sci.* **147**, 227–240 (1984).
17. Westley, M. S., Baragiola, R. A., Johnson, R. E. & Baratta, G. A. Photodesorption from low-temperature water ice in interstellar and circumsolar grains. *Nature* **373**, 405–407 (1995).
18. Simpson, W. C. *et al.* Dissociative electron attachment in nanoscale ice films: Temperature and morphology effects. *J. Chem. Phys.* **107**, 8668–8678 (1997).
19. Bednarek, J., Plonka, A., Hallbrucker, A., Mayer, E. & Symons, M. C. R. Hydroperoxyl radical generation by gamma-irradiation of glassy water at 77 K. *J. Am. Chem. Soc.* **118**, 9387–9390 (1996).
20. Taub, I. A. & Eiben, K. Transient solvated electron, hydroxyl, and hydroperoxy radicals in pulse-irradiated crystalline ice. *J. Chem. Phys.* **49**, 2499–2513 (1968).
21. Noll, K. S., Johnson, R. E., McGrath, M. A. & Caldwell, J. J. Detection of SO<sub>2</sub> on Callisto with the Hubble Space Telescope. *Geophys. Res. Lett.* **24**, 1139–1142 (1997).
22. Noll, K. S., Johnson, R. E., Lane, A. L., Domingue, D. L. & Weaver, H. A. Detection of ozone on Ganymede. *Science* **273**, 341–343 (1996).
23. Lacombe, S. *et al.* Electron-induced synthesis of ozone in a dioxygen matrix. *Phys. Rev. Lett.* **79**, 1146–1149 (1997).

**Acknowledgements.** We thank G. A. Kimmel for discussions. This work was supported by the US Department of Energy, Office of Basic Energy Sciences, Chemical Physics Program. Pacific Northwest National Laboratory is operated for the US Department of Energy by Battelle Memorial Institute.

Correspondence and requests for materials should be addressed to T.M.O. (e-mail: tm\_orlando@pnl.gov).

## Fluctuation-induced diffusive instabilities

David A. Kessler\* & Herbert Levine†

\* Minerva Center and Department of Physics, Bar-Ilan University, Ramat Gan, Israel

† Department of Physics, University of California, San Diego, La Jolla, California 92093-0319, USA

The formation of complex patterns in many non-equilibrium systems, ranging from solidifying alloys to multiphase flow<sup>1</sup>, nonlinear chemical reactions<sup>2</sup> and the growth of bacterial colonies<sup>3,4</sup>, involves the propagation of an interface that is unstable to diffusive motion. Most existing theoretical treatments of diffusive instabilities are based on mean-field approaches, such as the use of reaction–diffusion equations, that neglect the role of fluctuations. Here we show that finite fluctuations in particle number can be essential for such an instability to occur. We study, both analytically and with computer simulations, the planar

interface separating different species in the simple two-component reaction A + B → 2A (which can also serve as a simple model of bacterial growth in the presence of a nutrient). The interface displays markedly different dynamics within the reaction–diffusion treatment from that when fluctuations are taken into account. Our findings suggest that fluctuations can provide a new and general pattern-forming mechanism in non-equilibrium growth.

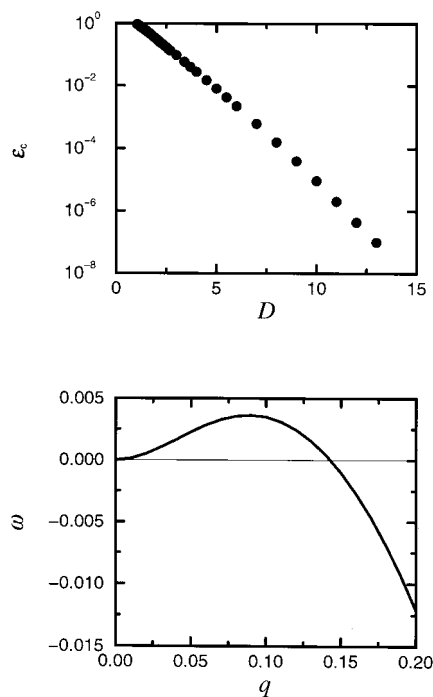
We consider a slightly modified version<sup>5</sup> of an “infection” model due to Blumen *et al.*<sup>6,7</sup>. There are two types of particles, infectious (A) and non-infected but susceptible (B). Both particle types move diffusively on a lattice, with no constraint on multiple occupancy, with differing rates D<sub>A</sub> and D<sub>B</sub>. A B particle which is co-located with an A particle has some probability per unit time of becoming infected, that is, changing to an A. At the mean-field level in which we ignore particle number fluctuations, the dynamics of this system is describable via the coupled reaction–diffusion equations:

$$\begin{aligned}\frac{\partial c_A}{\partial t} &= \nabla^2 c_A + c_A c_B \\ \frac{\partial c_B}{\partial t} &= D \nabla^2 c_B - c_A c_B\end{aligned}\quad (1)$$

Here we have scaled out the reaction rate and furthermore scaled space so as to eliminate D<sub>A</sub> in the first equation; D is defined as the ratio D<sub>B</sub>/D<sub>A</sub>. Formally, this equation can be obtained as the limit of the underlying Markov process when the average number of particles per site, N, goes to infinity. In the special case D = 1, this system reduces to the Fisher equation<sup>8</sup>.

Our interest here is in interface propagation, namely the process whereby introduction of some number of A particles at the left edge of a finite region filled with B particles at some initial concentration c<sub>B</sub><sup>0</sup> leads to the development of a moving front. The velocity of such a front in the reaction–diffusion approximation follows from the marginal stability principle<sup>9–11</sup>:

$$v_{MS} = 2\sqrt{c_B^0} \quad (2)$$



**Figure 1** Results of stability analysis. Top, dependence of the critical value  $\epsilon_c$  of the cut-off as a function of the diffusion-constant ratio  $D$ . Bottom, spectrum at  $D = 10$ ,  $\epsilon = 0.0014$ . See Methods for details.



In what follows we take  $c_B^0 = 1$  without loss of generality. It has recently been established<sup>5,12</sup> that, in models such as this one, the difference between the velocity of the microscopic process and that predicted by the above analysis exhibits the characteristic scaling form  $v_{MS} - v(N) \sim 1/\log^2 N$ . For example, a direct demonstration of this for the  $D = 1$  case was recently presented in Kessler *et al.*<sup>5</sup> Brunet and Derrida<sup>12</sup> have argued that this specific form can be understood as arising from an effective cut-off which needs to be placed into the continuum description, a cut-off arising from discrete particle number effects. This cut-off limits the reaction term to values of the A concentration greater than some small number  $\epsilon \approx 1/N$ , and leads to the above result, independent of the exact form of the cut-off. We note that similar ideas had been proposed earlier, in the contexts of diffusion-limited aggregation<sup>13</sup> and RNA virus evolution<sup>14</sup>, although a derivation of this deterministic cut-off approach (including for example, the exact correspondence between  $\epsilon$  and  $N$ ) from the master equation for the stochastic process is still lacking.

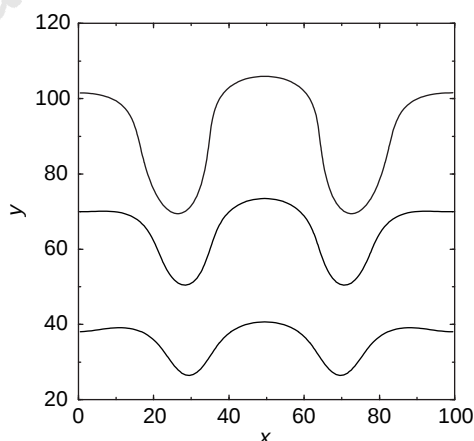
Although the velocity shift just described is quantitatively significant, the cut-off effect does not alter the qualitative form of the interface in a one-dimensional system. This changes dramatically when we go to higher dimensions. To show this, we focus on the stability of a planar interface in the modified continuum model obtained by setting the reaction term to zero if  $c_A$  is below  $\epsilon$ . The steady-state profile as a function of  $D$  and  $\epsilon$  is determined numerically, and the linear stability operator is defined in standard fashion by expanding the basic equations around this solution in the moving frame of reference. Because of translation invariance in the transverse directions, modes of this operator can be labelled by a transverse wavevector  $\mathbf{q}$  and the growth rate  $\omega$  is a function of the magnitude,  $q$ , of this vector. As the mode with  $\mathbf{q} = 0$  is merely a translation of the entire front, this must have growth rate zero.

Performing this calculation, we find that for a given diffusion ratio  $D$ , there is a critical value of the cut-off,  $\epsilon_c$ , such that the interface is unstable for all  $\epsilon > \epsilon_c$ . In Fig. 1a, we show a graph of  $\epsilon_c$  as a function of  $D$ . Equivalently, we may view the graph as showing the minimum value of  $D$  for which there is an instability at a given  $\epsilon$ . We see from the graph that this minimum  $D$  diverges as  $\epsilon \rightarrow 0$  as  $\ln(1/\epsilon)$ . This behaviour can be understood as follows. Diffusive instabilities require that the scale of the diffusing field  $D/v$  be sufficiently large compared to the width of the (linearly

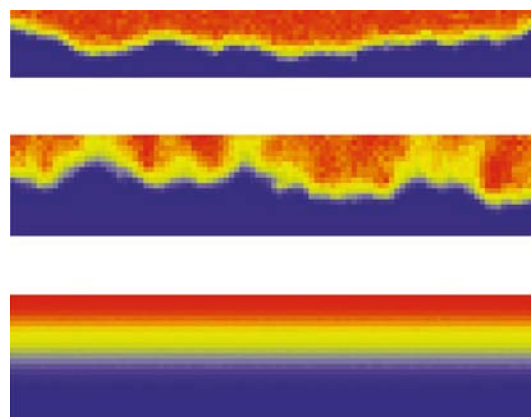
unstable) reaction zone. Here, the only thing limiting the spread of this zone is the finite cut-off. For small  $\epsilon$ , this width scales as  $\log(1/\epsilon)$ ; this was shown explicitly in Brunet and Derrida<sup>12</sup> for the cut-off Fisher equation and is easily extended to the present case. This explains why, in the absence of a cut-off, the interface is stable for all  $D$ . Thus we find that the naive reaction–diffusion approach misses the entire possibility that the interface can be unstable.

In Fig. 1b, we plot the full spectrum for a typical case ( $D = 10$ ,  $\epsilon = 0.0014$ ) inside the unstable region. We find that the instability begins at  $q = 0$  and extends to some finite  $q$ . This type of spectrum with its long wavelength instability is directly indicative of the diffusive nature of the instability; one could compare this graph to the Mullins–Sekerka<sup>15</sup> spectrum for alloy solidification, to take one example. In Fig. 2, we show results from a simulation of the cut-off reaction–diffusion system (for a different parameter set), showing the growth of initially small perturbations to the planar interface. Our theory predicts, therefore, that the non-equilibrium dynamics of our microscopic model exhibits a pattern-forming mechanism that is completely missed by the standard treatment. To verify this, we have performed a direct simulation of a two-dimensional system. We present in Fig. 3 the results of three computations, providing unambiguous evidence that this is indeed the case. The interface with  $D = 10$  and  $N = 1,000$  is diffusively unstable, as it supports large cellular structures with grooves where growth is screened out. Going to large  $N$  at fixed  $D$  eliminates the instability, as predicted by the analytic treatment. Finally, if we set  $D = 1$  and  $N = 1,000$ , the interface is again stable as predicted, but now shows significant fluctuations, as expected from the work of Kardar, Parisi and Zhang<sup>16</sup>.

One application of these ideas is to a series of experiments on unstable interfaces which arise during the nutrient-limited growth of bacterial colonies on agar<sup>3,4,17</sup>. The simplest model of such a process would contain (rapidly diffusing) nutrient particles (B) which are devoured by (slowly-diffusing) bacteria (A), thereby allowing the latter to reproduce. This model is very similar to Blumen kinetics, and can be expected to have exactly the same instability if one takes into account the cut-off effect; without the cut-off it is always stable (E. Ben-Jacob and I. Cohen, personal



**Figure 2** Results of a simulation of the reaction–diffusion equations with  $D = 10$  and  $\epsilon = 0.25$ . The equations are discretized on a grid of size  $100 \times 200$ , with lattice spacing one and periodic boundary conditions in  $x$ . The fields are updated using explicit time steps of  $\Delta t = 0.01$ . The lines shown in the graph are the  $c_A = 0.75$  contour lines for three different times, separated by 150 units. The graph depicts a region of size  $40 \times 100$  which spans the interface; the three different lines have been translated in  $y$  so as to better depict the shape change.



**Figure 3** Results of a simulation of the microscopic reaction model. The cases shown are the cases  $D = 10$  (top) and  $D = 1$  (middle), both for  $N = 1,000$  and a reaction rate constant  $k = 10^{-4}$ ; and  $D = 10$ ,  $N = 10^4$  with  $k = 10^{-5}$  (bottom). The graph shows the interfacial region for a box of width 100; the colour scale corresponds to the number of A particles (red is  $N_A \approx N$ , blue is  $N_A \approx 0$ ). Each site of a  $100 \times 200$  lattice has some number of A and B particles. Initially, A particles are introduced at one edge, whereas B particles exist everywhere, with an average density of  $N$ . In each time step, particles can diffuse and same-site particles can react. The data shown is a typical snapshot of the interface once a statistical steady-state has been established.

communication). It is not yet clear whether this idea can quantitatively account for the instability, or whether a more complex model (possibly incorporating nonlinear diffusion<sup>18</sup>), which can give rise to instabilities even at the mean-field level, is more appropriate.

From the perspective of the general study of non-equilibrium physics, our finding are quite unusual. The fact that inevitable fluctuation effects in a finite density process can lead to such a drastic change in its large-scale properties is unexpected. This result complicates any attempt to use the calculational machinery of equilibrium statistical mechanics to study these systems, as these methods start from the point of view that fluctuation effects (in the form of additive thermal noise) become important only near a critical point. Here (and in the related process of diffusion-limited aggregation<sup>19</sup>), the fluctuations are due to discreteness (that is, they are shot-noise-like), enter multiplicatively and can be important even far from any critical scaling region. Our results indicate that a more sophisticated methodology which can account for the zeroth-order relevance of discreteness is a necessary starting point for analysing a wide class of interesting and important dynamical processes. □

## Methods

The steady-state equation is

$$0 = v c_A' + c_A'' + c_A c_B$$

$$0 = v c_B' + D c_B'' - c_A c_B$$

with  $c_A(\infty) = 0$ ,  $c_B(\infty) = 1$  and ' means  $d/dx$  in the moving frame. If we denote by  $x_c$  the position at which  $c_A = \epsilon$ , we find (by matching to the region  $x > x_c$ ;

$$c_A(x_c) = \epsilon \quad c_A'(x_c) = -\epsilon v$$

$$c_B(x_c) = 1 - \delta_1 \quad c_B'(x_c) = \delta_1 v/D$$

where  $\delta_1$  is as yet unknown. At large negative  $x$ , the most general asymptotic behaviour is;

$$c_A(x) \approx 1 + \delta_2 - \delta_3 e^{kx}$$

$$c_B(x) \approx \delta_3 (vk + k^2) e^{kx}$$

with  $k = -v/2D + [(v/2D)^2 + (1/D)]^{1/2}$ . We integrate from both directions to  $x = 0$  and impose  $c_A(0) = 0.5$  so as to fix the translation invariance. These two conditions plus three continuity equations suffice to determine the five unknowns  $x_c$ ,  $v$  and the  $\delta$ s. In fact,  $\delta_2$  should be exactly zero due to a global conservation equation, and this can be used as a check on the accuracy of the numerical integration.

Once the steady-state is determined, we use a similar procedure for the stability spectrum. The equations we solve are, for  $x < x_c$ ;

$$(\omega + q^2)\psi_A = v\psi_A' + \psi_A'' + \psi_A c_B^{(0)} + \psi_B c_A^{(0)}$$

$$(\omega + Dq^2)\psi_B = v\psi_B' + D\psi_B'' - \psi_A c_B^{(0)} - \psi_B c_A^{(0)}$$

where  $c_{A,B} = c_{A,B}^{(0)} + \psi_{A,B} \exp(iq \cdot y + \omega t)$ . At large negative  $x$ , the asymptotic behaviour of the most general allowed solution for  $\text{Re } \omega > 0$  consists of;

$$\psi_A(x) \approx A_1 e^{k_1 x} + A_2 e^{k_2 x}$$

$$\psi_B(x) \approx -A_1 (vk_1 + k_1^2 - q^2 - \omega) e^{k_1 x}$$

where

$$k_1 = -\frac{v}{2D} + \sqrt{\left(\frac{v}{2D}\right)^2 + \frac{1 + \omega + Dq^2}{D}}$$

$$k_2 = -\frac{v}{2} + \sqrt{\left(\frac{v}{2}\right)^2 + \omega + q^2}$$

The situation at  $x_c$  needs to be handled carefully, as one must account for the fact that the perturbation in  $c_A$  will in general shift the position at which the cut-off sets in. This leads to the boundary conditions;

$$\psi_A(x_c) = A_4 \quad \psi_A'(x_c) = A_4(k_4 + c_B^{(0)}(x_c)/v)$$

$$\psi_B(x_c) = A_3 \quad \psi_B'(x_c) = A_3 k_3 - A_4 c_B^{(0)}(x_c)/Dv$$

where now

$$k_3 = -\frac{v}{2D} - \sqrt{\left(\frac{v}{2D}\right)^2 + \frac{\omega + Dq^2}{D}}$$

$$k_4 = -\frac{v}{2} - \sqrt{\left(\frac{v}{2}\right)^2 + \omega + q^2}$$

Imposing continuity conditions at  $x = 0$  leads to a set of four homogeneous linear equation for the four  $A$ s. The vanishing of the determinant of this system then determines the eigenvalue  $\omega$ . This last step is implemented by a standard nonlinear equation solver. As a check on the accuracy of our procedure, we can solve for  $\omega(q = 0)$  which should be precisely zero as it is just the translation mode. This number is typically of the order of  $10^{-6}$  for our calculations.

Received 12 March; accepted 18 May 1998.

- Kessler, D., Koplik, J. & Levine, H. Pattern selection in fingered growth phenomena. *Adv. Phys.* **37**, 255–339 (1988).
- Horvath, D. & Toth, A. Diffusion-driven front instabilities in the chlorite-tetrathionate reaction. *J. Chem. Phys.* **108**, 1447–1451 (1998).
- Matsushita, M. & Fujikawa, H. H. Diffusion-limited growth in bacterial colony formation. *Physica A* **168**, 498–506 (1990).
- Ben-Jacob, E. *et al.* Generic modelling of cooperative growth patterns in bacterial colonies. *Nature* **368**, 46–49 (1994).
- Kessler, D., Ner, Z. & Sander, L. M. Front propagation: precursors, cutoffs and structural stability. *Phys. Rev. E* **58**, 107–114 (1998).
- Mai, J., Sokolov, I. M. & Blumen, A. Front propagation and local ordering in one-dimensional irreversible autocatalytic reactions. *Phys. Rev. Lett.* **77**, 4462–4465 (1996).
- Mai, J., Sokolov, I. M., Kuzovkov, V. N. & Blumen, A. Front form and velocity in a one-dimensional autocatalytic A+B → 2A reaction. *Phys. Rev. E* **56**, 4130–4134 (1997).
- Fisher, R. A. The wave of advance of advantageous genes. *Ann. Eugen.* **7**, 355–369 (1937).
- Ben-Jacob, E., Brand, H., Dee, G., Kramer, L. & Langer, J. S. Pattern propagation in nonlinear dissipative systems. *Physica D* **14**, 348–364 (1985).
- van Saarloos, W. Dynamical velocity selection: marginal stability. *Phys. Rev. Lett.* **58**, 2571–2574 (1987).
- Paquette, G. C., Chen, L.-Y., Goldenfeld, N. & Oono, Y. Structural stability and renormalization group for propagating fronts. *Phys. Rev. Lett.* **72**, 76–79 (1994).
- Brunet, E. & Derrida, B. Shift in the velocity of a front due to a cutoff. *Phys. Rev. E* **65**, 2597–2604 (1997).
- Brener, E., Levine, H. & Tu, Y. Mean-field theory for diffusion-limited aggregation in low dimensions. *Phys. Rev. Lett.* **66**, 1978–1981 (1991).
- Tsimring, L. S., Levine, H. & Kessler, D. A. RNA virus evolution via a fitness-space model. *Phys. Rev. Lett.* **76**, 4440–4443 (1996).
- Mullins, W. W. & Sekerka, R. F. Stability of a planar interface during solidification of a dilute binary alloy. *J. Appl. Phys.* **35**, 444–451 (1964).
- Kardar, M., Parisi, G. & Zhang, Y. C. Dynamic scaling of growing interfaces. *Phys. Rev. Lett.* **56**, 889–892 (1986).
- Ben-Jacob, E. *et al.* Response of bacterial colonies to imposed anisotropy. *Phys. Rev. E* **53**, 1835–1843 (1996).
- Newman, W. I. The long-time behavior of solutions to a nonlinear diffusion problem in population dynamics and combustion. *J. Theor. Biol.* **104**, 473–484 (1983).
- Witten, T. A. & Sander, L. M. Diffusion-limited aggregation, a kinetic critical phenomenon. *Phys. Rev. Lett.* **47**, 1400–1403 (1981).

**Acknowledgements.** We thank L. Sander for discussions.

Correspondence and requests for materials should be addressed to H.L. (e-mail: hlevine@ucsd.edu).

## Time-reversal symmetry-breaking superconductivity in Sr<sub>2</sub>RuO<sub>4</sub>

G. M. Luke\*, Y. Fudamoto\*, K. M. Kojima\*, M. I. Larkin\*, J. Merrin\*, B. Nachumi\*, Y. J. Uemura\*, Y. Maeno†, Z. Q. Mao‡, Y. Mori†, H. Nakamura‡ & M. Sigrist§

\* Department of Physics, Columbia University, New York, New York 10027, USA

† Department of Physics and § Yukawa Institute for Theoretical Physics, Kyoto University, Kyoto 606-8502, Japan

‡ Department of Material Science and Engineering, Kyoto University, Kyoto 606-8501, Japan

Although the properties of most superconducting materials are well described by the theory<sup>1</sup> of Bardeen, Cooper and Schrieffer (BCS), considerable effort has been devoted to the search for exotic superconducting systems in which BCS theory does not apply. The transition to the superconducting state in conventional BCS superconductors involves the breaking of gauge symmetry

only, whereby the wavefunction describing the Cooper pairs—the paired electron states responsible for superconductivity—adopt a definite phase. In contrast, a signature of an unconventional superconducting state is the breaking of additional symmetries<sup>2</sup>, which can lead to anisotropic pairing (such as the ‘*d*-wave’ symmetry observed in the copper oxide superconductors) and the presence of multiple superconducting phases (as seen in UPt<sub>3</sub> and analogous behaviour in superfluid <sup>3</sup>He; refs 3–5). Here we report muon spin-relaxation measurements on the superconductor Sr<sub>2</sub>RuO<sub>4</sub> that reveal the spontaneous appearance of an internal magnetic field below the transition temperature: the appearance of such a field indicates that the superconducting state in this material is characterized by the breaking of time-reversal symmetry. These results, combined with other symmetry considerations, suggest that superconductivity in Sr<sub>2</sub>RuO<sub>4</sub> is of ‘*p*-wave’ (odd-parity) type, analogous to superfluid <sup>3</sup>He.

Sr<sub>2</sub>RuO<sub>4</sub>, which is isostructural with the high-*T*<sub>c</sub> copper oxide La<sub>1.85</sub>Sr<sub>0.15</sub>CuO<sub>4</sub>, is to date the only known layered perovskite superconductor which does not contain copper (here *T*<sub>c</sub> is the superconducting transition temperature). Although first synthesized in the 1950s (ref. 6), its superconductivity was only found in 1994 (ref. 7); *T*<sub>c</sub>s of early samples were ~0.7 K but have increased to *T*<sub>c</sub> = 1.5 K in recent high-quality single crystals<sup>8</sup>. Despite its low transition temperature, Sr<sub>2</sub>RuO<sub>4</sub> is of great interest as there is growing evidence for an unconventional superconducting state in this compound. In this system, strong correlation effects enhance the effective mass seen in quantum oscillation<sup>9</sup> and Pauli spin susceptibility measurements, in the same way as in <sup>3</sup>He (ref. 10). Combining this feature with the expected tendency of Sr<sub>2</sub>RuO<sub>4</sub> to display ferromagnetic spin fluctuations, Rice and Sigrist<sup>11</sup>, and later Baskaran<sup>12</sup>, argued that the pairing in Sr<sub>2</sub>RuO<sub>4</sub> could be of odd-parity (spin triplet) type.

The strong suppression of the superconducting *T*<sub>c</sub> by even non-magnetic impurities suggests non-*s*-wave pairing<sup>8</sup>. Specific heat<sup>13</sup> and NMR 1/*T*<sub>1</sub> (ref. 14) measurements indicate the presence of a large residual density of states (RDOS) at low temperatures (well within the superconducting state); in high-quality samples, this RDOS seems to approach half of the normal state value as *T* → 0. Several authors<sup>15,16</sup> have proposed so-called non-unitary *p*-wave superconducting states for Sr<sub>2</sub>RuO<sub>4</sub> to account for this RDOS as well as the absence of a Hebel–Slichter peak in NMR measurements<sup>14</sup>. A finite RDOS is not a unique signature of unconventional superconductivity; for example, it is observed in so-called gapless superconductors with isotropic *s*-wave pairing as a result of impurity scattering (although this is unlikely in the specific case of Sr<sub>2</sub>RuO<sub>4</sub> where the finite RDOS apparently remains in the cleanest samples). It could also be explained with a multi-band hypothesis<sup>17</sup> where different gaps are associated with two types of bands. Further studies are therefore required for a definitive determination of the pairing symmetry in Sr<sub>2</sub>RuO<sub>4</sub>.

One aspect of the pairing symmetry, the breaking of time-reversal symmetry (TRS), can be probed directly. If the superconducting state has a degenerate representation (as is possible for some triplet superconducting states) then TRS can be broken, whereas it cannot be broken for non-degenerate representations (the case for all singlet states). When spin–orbit coupling is small, the pair wavefunction can be written as a product of the orbital part and the spin part. Non-zero angular momentum of either the orbital or the spin part could result in TRS breaking, although there are many such cases with conserved TRS, such as *d*-wave states in the high-*T*<sub>c</sub> copper oxides. TRS breaking is also possible in the case of strong spin–orbit coupling. In general, pairing states can be further classified in terms of gap functions attached to irreducible representations of a point group for a given crystal lattice symmetry of the system<sup>2</sup>. TRS is broken for some of the unitary states and for all of the non-unitary states. For example, the B phase of <sup>3</sup>He has a unitary state which conserves TRS, the A phase is unitary but breaks

TRS, and so does the non-unitary A<sub>1</sub> phase which is stable under an applied external magnetic field. In contrast, conventional superconductors (including the BCS state) are unitary and conserve TRS.

For states with broken TRS, originating from either spin or orbital moments, a spontaneous internal magnetic field can appear below *T*<sub>c</sub>. One can expect a finite hyperfine field at a magnetic probe for the case of spin moments, while (for both the spin and orbital cases) spontaneous supercurrents in the vicinity of inhomogeneities in the order parameter would create a magnetic field near impurities, surfaces, and/or domain walls between the two degenerate superconducting phases<sup>2</sup>.

Muon spin relaxation<sup>18</sup> (μSR) is especially powerful for detecting such effects of TRS breaking in exotic superconductors, because internal magnetic fields as small as 0.1 G—corresponding to ordered moments ~0.01 μ<sub>B</sub>—are readily detected in measurements without applying external magnetic fields. Furthermore, the presence of a positively charged μ<sup>+</sup> particle could distort the electrostatic potential and perturb the superconducting order parameter, inducing magnetic fields. μSR was previously used in a search for a spontaneous internal field predicted by an orbital state which breaks TRS in the ‘anyon’ hypothesis proposed for high-*T*<sub>c</sub> copper oxides. No evidence for broken TRS was observed<sup>19</sup> in that case. However, a spontaneous internal field was observed by μSR in the so-called B phase of UPt<sub>3</sub> (ref. 20), and in (U,Th)Be<sub>13</sub> (ref. 21). A number of authors have proposed states (both unitary and non-unitary) which break TRS for the B phase of UPt<sub>3</sub>, consistent with this observation<sup>22</sup>.

In a μSR experiment, 100% spin-polarized positive muons are implanted one at a time into a specimen. After quickly coming to rest (at a typical depth of ~0.1 mm), the muon spins evolve in the local magnetic environment. The muon subsequently decays (τ<sub>μ</sub> = 2.2 μs), emitting a positron preferentially in the direction of the muon spin at the time of decay. By accumulating time histograms of many (10<sup>7</sup>) such positrons, one may deduce the muon polarization function as a function of time after implantation. In a zero field (ZF-μSR) experiment, positron detectors are positioned 180° apart, parallel and antiparallel to the initial polarization direction. Apart from normalization factors which depend on the experimental geometry and the properties of muon decay, the asymmetry signal—which is defined as the difference divided by the sum of the count rates of the two opposing detectors—directly represents the muon polarization at the time of decay.

In the absence of magnetic order, the spin polarization is relaxed by static randomly orientated nuclear dipole moments and is well described by the Kubo–Toyabe function<sup>23</sup>:

$$P_{\mu} = \frac{1}{3} + \frac{2}{3}(1 - \Delta^2 t^2) \exp[-\frac{1}{2}\Delta^2 t^2] \quad (1)$$

where Δ/γ<sub>μ</sub> is the width of the local field distribution and γ<sub>μ</sub> = 0.085 μs<sup>-1</sup> G<sup>-1</sup> is the muon gyromagnetic ratio. In the magnetically ordered state, the μSR spectrum will exhibit precession (where the frequency is proportional to the ordered moment) if the internal field is sufficiently uniform. If the spontaneous field is weak, or there is a broad distribution of local fields (as in, for example, a spin glass) then one observes an increase in the relaxation of *P*<sub>μ</sub>, where the increase in relaxation corresponds to an order parameter.

Two single crystals of Sr<sub>2</sub>RuO<sub>4</sub> were grown at Kyoto University using a floating-zone method with an infrared image furnace. The feed rod, containing 10% excess Ru serving as flux, was melted in a 10%-O<sub>2</sub>/90%-Ar gas mixture with a total pressure of 2 bar (sample A) or 3 bar (sample B). The crystal growth was performed at speeds of 50 mm h<sup>-1</sup> (sample A) and 40 mm h<sup>-1</sup> (sample B). The superconducting *T*<sub>c</sub>s were 1.478 K and 1.453 K with widths of 48 mK and 40 mK, respectively, where *T*<sub>c</sub> is defined as the sharp rise in the dissipative component of the a.c.-susceptibility and the width is its full-width at half-maximum. Sample B was annealed in air at 1,500 °C for 72 h to reduce defects, whereas sample A was not annealed. The fact that the two samples had high *T*<sub>c</sub>s (for Sr<sub>2</sub>RuO<sub>4</sub>) is indicative of their high quality and demonstrates that

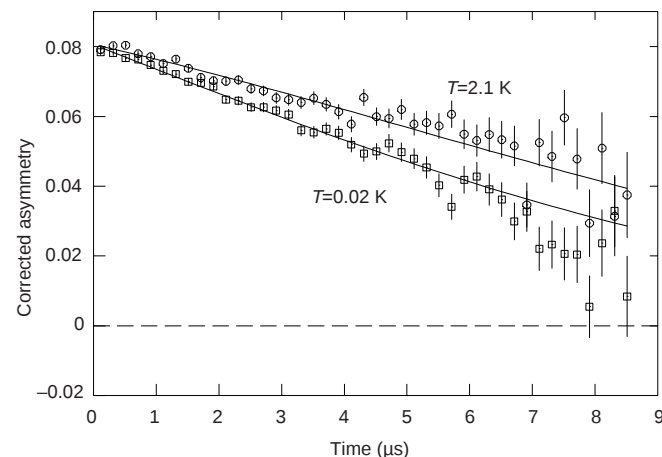


any impurities must be extremely dilute<sup>8</sup>. Sample A was cleaved and arranged in a mosaic so that the *c*-axis was normal to the sample's planar surface and parallel to the initial muon polarization, whereas sample B was cut using a diamond saw such that the *c*-axis lay in the plane of the sample. Each specimen covered an area of  $\sim 1 \text{ cm}^2$ .

We performed ZF- $\mu$ SR measurements on the M15 surface muon channel at TRIUMF using an Oxford 400 top-loading dilution refrigerator. Sample A was mounted on a silver backing, whereas sample B was mounted on high-purity GaAs. Both backing materials give a temperature-independent and essentially non-relaxing  $\mu$ SR signal in zero-field measurements. Stray magnetic fields at the sample position were reduced to less than 0.1 G in all directions. Temperatures were measured with a calibrated carbon resistor mounted on the mixing chamber, as well as with a  $\text{RuO}_2$  resistor adjacent to the sample.

In Fig. 1 we show ZF- $\mu$ SR asymmetry spectra measured at  $T = 2.1 \text{ K}$  and  $T = 0.02 \text{ K}$  for sample B, where  $P_\mu \perp c$ . Spectra for sample A ( $P_\mu \parallel c$ ) are qualitatively similar. From Fig. 1 we see that the relaxation rate is quite small both above and below  $T_c = 1.45 \text{ K}$ , but that there is greater relaxation at lower temperature. This increased zero-field relaxation could be caused by either (quasi-)static or fluctuating magnetic fields. To distinguish between these two possibilities, we performed measurements in a weak longitudinal field  $B_{\text{LF}} = 50 \text{ G}$ . Relaxation due to static fields will be decoupled by the application of an external field greater than several times the characteristic internal field, whereas decoupling of dynamic relaxation requires much greater applied fields<sup>18</sup>. We see in the lower panel of Fig. 2 that the relaxation in  $B_{\text{LF}} = 50 \text{ G}$  is the same above and below  $T_c$ , which indicates that the zero-field relaxation must therefore be due to spontaneous fields which are static on the micro-second timescale.

We tried several functional forms for the additional relaxation below  $T_c$ : fits to precessing (cosine) or gaussian forms were essentially equivalent to each other, but were significantly worse than an exponential in fitting the data. Fitting the spectra to the product of equation (1) (using common values of  $\Delta = 0.02$ ,  $0.06 \mu\text{s}^{-1}$  for samples A and B, respectively, for all temperatures to account for the nuclear dipole fields) and an exponential  $\exp(-\Lambda t)$  to characterize the additional relaxation due to the spontaneous magnetic field, we plot  $\Lambda$  versus  $T$  in Fig. 2 for samples A ( $P_\mu \parallel c$ ) and B ( $P_\mu \perp c$ ). Also shown are the superconducting  $T_c$ s as determined by a.c.-susceptibility. We see a clear increase in  $\Lambda$  with an onset temperature around  $T_c = 1.45 \text{ K}$  for both orientations indicating the presence of a spontaneous magnetic field appearing within the superconducting state. In principle, this spontaneous field could originate either from a TRS-breaking superconducting



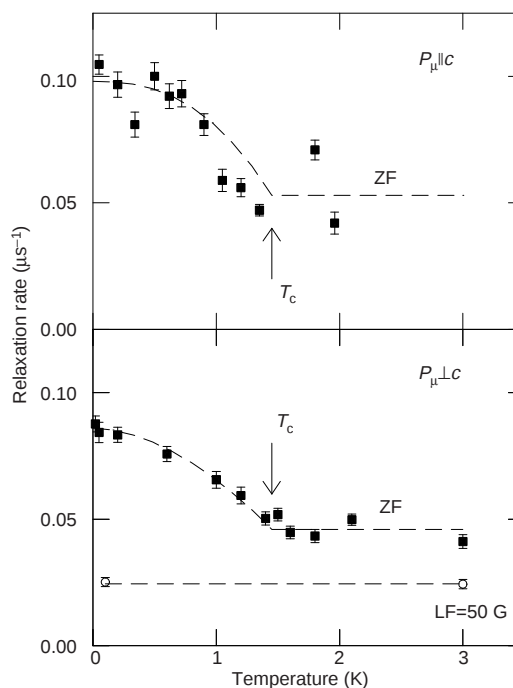
**Figure 1** Zero-field  $\mu$ SR spectra measured with  $P_\mu \perp c$  in  $\text{Sr}_2\text{RuO}_4$  at  $T = 2.1 \text{ K}$  (circles) and  $T = 0.02 \text{ K}$  (squares). Curves are fits to the product of equation (1) and  $\exp(-\Lambda t)$  as described in the text.

state or with a purely magnetic state which coincidentally appears near  $T_c$ . We performed additional measurements on a third sample with a reduced  $T_c$  of  $0.9 \text{ K}$ , and found that the spontaneous field occurred at that reduced  $T_c$ . Thus we conclude that because this field is so intimately connected with superconductivity, its existence provides direct evidence for a TRS-breaking superconducting state in  $\text{Sr}_2\text{RuO}_4$ .

The increased relaxation in the superconducting state is exponential in character. If there were a unique field at the muon site we would observe a precession signal, whereas a dense collection of randomly orientated field sources would result in a Kubo–Toyabe relaxation function. Both the precession and Kubo–Toyabe signals start with a gaussian form and in the limit of weak characteristic fields are essentially indistinguishable. The exponential form which we observe in  $\text{Sr}_2\text{RuO}_4$  indicates a broad distribution of fields arising from a dilute distribution of sources<sup>23</sup>. This is consistent with the source of the internal field being supercurrents associated with variations in the superconducting order parameter around dilute impurities and domain walls. We see enhanced relaxation in both samples A and B. From this we conclude that the local field cannot lie purely along the *c*-axis (this would give no relaxation from sample A). However, any orientation of the local field which has a significant component in the basal plane is consistent with our data.

The increase in the exponential relaxation below  $T_c$  is  $\sim 0.04 \mu\text{s}^{-1}$ , which corresponds to a characteristic field strength  $\Lambda/\gamma_\mu = 0.5 \text{ G}$ . This is about the same as we observed in the B phase of  $\text{UPt}_3$  (ref. 20) and is in line with theoretical predictions for that material. No theoretical estimates of the characteristic field strength in  $\text{Sr}_2\text{RuO}_4$  are yet available; however, we expect them to be comparable to those in  $\text{UPt}_3$  as the fields should arise from a similar mechanism.

Several authors have considered the allowed symmetries of pairing states for the specific case of tetragonal symmetry appropriate for  $\text{Sr}_2\text{RuO}_4$  (refs 2, 11, 15, 16). Spin-singlet pairing leads exclusively to non-degenerate states and the breakdown of TRS would, in general, require additional phase transitions admixing other pairing channels within the superconducting phase, for which there is no experimental indication so far. On the other hand, one can show that in the spin-triplet pairing channel, TRS-breaking



**Figure 2** Zero-field (ZF) relaxation rate  $\Lambda$  for the initial muon spin polarization  $\parallel c$  (top) and  $\perp c$  (bottom).  $T_c$  from a.c.-susceptibility indicated by arrows. Circles in bottom figure give relaxation rate in  $B_{\text{LF}} = 50 \text{ G}$ . Curves are guides to the eye.

states appear naturally at the onset of superconductivity, consistent with our experiment. On the basis of these symmetry arguments we conclude that the present experiment provides strong evidence for Cooper pairing with spin-triplet ( $p$ -wave) symmetry: a superconducting analogue of the A or A<sub>1</sub> phases of superfluid <sup>3</sup>He. The distinction between unitary and non-unitary states in Sr<sub>2</sub>RuO<sub>4</sub>, however, cannot be made with the present results, and has to wait for further studies by other means. □

Received 27 March; accepted 30 June 1998.

1. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. Theory of superconductivity. *Phys. Rev.* **108**, 1175–1204 (1957).
2. Sigrist, M. & Ueda, K. Phenomenological theory of unconventional superconductivity. *Rev. Mod. Phys.* **63**, 239–311 (1991).
3. Lee, D. M. The extraordinary phases of liquid <sup>3</sup>He. *Rev. Mod. Phys.* **69**, 645–665 (1997).
4. Osheroff, D. D. Superfluidity in <sup>3</sup>He: discovery and understanding. *Rev. Mod. Phys.* **69**, 667–681 (1997).
5. Richardson, R. C. The Pomeranchuk effect. *Rev. Mod. Phys.* **69**, 683–690 (1997).
6. Randall, J. J. & Ward, R. J. The preparation of some ternary oxides of the platinum metals. *J. Am. Chem. Soc.* **81**, 2629–2631 (1959).
7. Maeno, Y. *et al.* Superconductivity in a layered perovskite without copper. *Nature* **372**, 532–534 (1994).
8. Mackenzie, A. P. *et al.* Extremely strong dependence of superconductivity on disorder in Sr<sub>2</sub>RuO<sub>4</sub>. *Phys. Rev. Lett.* **80**, 161–164 (1998).
9. Mackenzie, A. P. *et al.* Quantum oscillations in the layered perovskite superconductor Sr<sub>2</sub>RuO<sub>4</sub>. *Phys. Rev. Lett.* **76**, 3786–3789 (1996).
10. Vollhardt, D. & Wölfle, P. *The Superfluid Phases of Helium 3* (Taylor & Francis, London, 1990).
11. Rice, T. M. & Sigrist, M. Sr<sub>2</sub>RuO<sub>4</sub>: an electronic analogue of <sup>3</sup>He? *J. Phys. Condens. Matter* **7**, 643–648 (1995).
12. Baskaran, G. Why is Sr<sub>2</sub>RuO<sub>4</sub> not a high T<sub>c</sub> superconductor? Electron correlation, Hund's coupling and  $p$ -wave instability. *Physica B* **223 & 224**, 490–495 (1996).
13. Nishizaki, S., Maeno, Y., Farner, S., Ikeda, S. & Fujita, T. Pairing symmetry of superconducting Sr<sub>2</sub>RuO<sub>4</sub> from specific heat measurements. *Physica C* **282–287**, 1413–1414 (1997).
14. Ishida, K. *et al.* Anisotropic pairing in superconducting Sr<sub>2</sub>RuO<sub>4</sub>: Ru NMR and NQR studies. *Phys. Rev. B* **56**, R505–R508 (1997).
15. Sigrist, M. & Zhitomirsky, M. E. Pairing symmetry of the superconductor Sr<sub>2</sub>RuO<sub>4</sub>. *J. Phys. Soc. Jpn* **65**, 3452–3455 (1996).
16. Machida, K., Ozaki, M. & Ohmi, T. Odd-parity pairing superconductivity under tetragonal symmetry—possible application to Sr<sub>2</sub>RuO<sub>4</sub>. *J. Phys. Soc. Jpn* **65**, 3720–3723 (1996).
17. Agterberg, D. F., Rice, T. M. & Sigrist, M. Orbital dependent superconductivity in Sr<sub>2</sub>RuO<sub>4</sub>. *Phys. Rev. Lett.* **78**, 3374–3377 (1997).
18. Brewer, J. H. in *Encyclopedia of Applied Physics* Vol. 11 (ed. Trigg, G. L.) 23 (VCH, New York, 1994).
19. Kiefl, R. F. *et al.* Search for anomalous internal magnetic fields in high-T<sub>c</sub> superconductors as evidence for broken time-reversal symmetry. *Phys. Rev. Lett.* **64**, 2082–2085 (1990).
20. Luke, G. M. *et al.* Muon spin relaxation in UPt<sub>3</sub>. *Phys. Rev. Lett.* **71**, 1466–1469 (1993).
21. Heffner, R. H. *et al.* New phase diagram for (U<sub>1-x</sub>Th)<sub>x</sub>Be<sub>13</sub>: a muon-spin-resonance and H<sub>rf</sub> study. *Phys. Rev. Lett.* **65**, 2816–2819 (1990).
22. Heffner, R. H. & Norman, M. R. Heavy fermion superconductivity. *Comments Condens. Matter Phys.* **17**, 361–408 (1996).
23. Uemura, Y. J., Yamazaki, T., Harshman, D. R., Senba, M. & Ansaldo, E. J. Muon-spin relaxation in AuFe and CuMn spin glasses. *Phys. Rev. B* **31**, 546–563 (1985).

**Acknowledgements.** We thank K. Machida, D. Agterberg and E. M. Forgan for discussions. Research at Columbia was supported by NSF and NEDO. Y.M. and Z.Q.M. thank CREST of the Japan Science and Technology Corporation for its support.

Correspondence and requests for materials should be addressed to G.M.L. (e-mail: luke@phys.columbia.edu).

## Weekly cycles of air pollutants, precipitation and tropical cyclones in the coastal NW Atlantic region

Randall S. Cervený & Robert C. Balling Jr

Office of Climatology and Department of Geography, Arizona State University, Tempe, Arizona 85287-1508, USA

Direct human influences on climate have been detected at local scales, such as urban temperature increases and precipitation enhancement<sup>1–3</sup>, and at global scales<sup>4,5</sup>. A possible indication of an anthropogenic effect on regional climate is by identification of equivalent weekly cycles in climate and pollution variables. Weekly cycles have been observed in both global surface temperature<sup>6</sup> and local pollution<sup>7</sup> data sets. Here we describe

statistical analyses that reveal weekly cycles in three independent regional-scale coastal Atlantic data sets: lower-troposphere pollution, precipitation and tropical cyclones. Three atmospheric monitoring stations record minimum concentrations of ozone and carbon monoxide early in the week, while highest concentrations are observed later in the week. This air-pollution cycle corresponds to observed weekly variability in regional rainfall and tropical cyclones. Specifically, satellite-based precipitation estimates indicate that near-coastal ocean areas receive significantly more precipitation at weekends than on weekdays. Near-coastal tropical cyclones have, on average, significantly weaker surface winds, higher surface pressure and higher frequency at weekends. Although our statistical findings limit the identification of cause–effect relationships, we advance the hypothesis that the thermal influence of pollution-derived aerosols on storms may drive these weekly climate cycles.

High concentrations of anthropogenic aerosols have been identified over the North Atlantic Ocean<sup>8–10</sup>: a substantial amount of these pollutants are advected from the urbanized eastern seaboard of North America<sup>8,11</sup>. We note that urban centres have a strong weekly pollution cycle<sup>7,12</sup>. This variability, nicknamed the “Sunday effect”<sup>7</sup>, is characterized by high late-week pollution levels as opposed to the early week<sup>13</sup> and, although still debated, is probably the result of car driving<sup>13</sup>. United States and Canadian inventories confirm higher weekday emissions as opposed to those of the weekend<sup>14</sup>.

If east-coast metropolitan areas display weekly pollution cycles and advection of that pollution is occurring, corresponding pollution cycles should be evident at downwind monitoring stations. To test this hypothesis, we used a 965-day (July 1991–January 1995) air-quality data set for Sable island (Fig. 1) as an indicator of pollution transport into the North Atlantic Ocean<sup>8</sup>. Concentrations (in p.p.b.v.) of two pollutants were monitored: ozone (O<sub>3</sub>) and carbon monoxide (CO); a tracer for anthropogenic pollution<sup>15</sup>.

Because certain statistical techniques require a normal distribution, all data were tested for significant deviations from normality by computing standardized skewness and kurtosis coefficients<sup>16,17</sup>. Atlantic pollution data sets met without transformation our significance criterion of  $\alpha = 0.05$  which is used for all tests in this study. Barlett's test of variance dispersion reveals no significant difference in variance by day of the week. Displayed analysis is limited to data from April to October; research indicates that this period is most associated with ozone advection for the northern North Atlantic<sup>8</sup>. Recognizing that some data may have significant autocorrelation, we removed first-order autocorrelation from all data and found insignificant effects on our results. First-order autocorrelation results were also not significantly influenced by the few intermittent runs of zeroes in the precipitation dataset. We also adjusted the sample size for each analysis to reflect the potential influence of first-order autocorrelation, and again found no fundamental change in our results.

A seven-day cycle is evident in both O<sub>3</sub> and CO time series, with the late week (Thursday–Friday) experiencing highest values of pollutants and with lowest values associated with the early week (Sunday–Tuesday). After the two pollutants were standardized and combined into a single variable (Fig. 2a), a series of statistical analyses confirm weekly cyclicity in the CO + O<sub>3</sub> variable. First, a Student's  $t$ -test indicated a significant difference in daily mean concentrations between Monday and Thursday. Additional testing (with  $\alpha = 0.0024$ ) using the assumption of completely independent daily data (to avoid the problem of multiplicity) still confirms the significant difference between the two extremes of the week. Second, spectral analysis identifies peaks in spectral density at seven days in CO, O<sub>3</sub> and CO + O<sub>3</sub> concentrations. Third, when the pollution data are categorized into 127 weekly periods, a 7-day sine wave (first harmonic) explains an average 50% of the variance in each weekly period. Fourth, shorter (~61 days, July–September 1991) data sets for Canadian Coast Guard lighthouses at Seal island, Nova Scotia

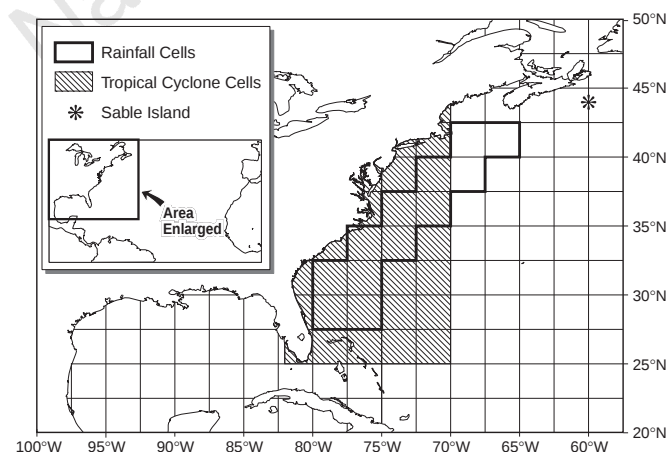
and Cape Race, Newfoundland<sup>8</sup>, as well as Sable island, also indicate significant differences in CO and O<sub>3</sub> between early- and late-week observations. Fifth, an independent pollution data set<sup>18</sup> for Bermuda (where, owing to circulation, North American pollution advection creates early spring maxima in CO and O<sub>3</sub>; ref. 11) shows a similar, but out of phase, weekly cycle in March and April with a mid-week pollution maximum. This phase shift in pollution transport between Sable island and Bermuda may result from changes in seasonal circulation strength and location.

Weekly cycles are associated with human activities rather than with natural phenomena<sup>6</sup>. No meteorological mechanism has a consistent seven-day return period. Consequently, identification of seven-day cycles in climate variables, coupled with the pollution cycle above, suggests direct anthropogenic forcing of regional climate. Investigation of such linkages is not new; examination of corresponding weekly cycles in rainfall and pollution date back to 1929<sup>19</sup>. For the present investigation, two independent climate data bases are examined. The first is a satellite record of daily precipitation over the world's oceans<sup>20</sup>.

Daily oceanic precipitation values are available from Microwave Sounding Unit (MSU) channels 1, 2 and 3 as gathered by seven separate TIROS-N satellites<sup>20</sup> for 1 January 1979 to 31 March 1995. For this study, data were analysed for all oceanic 2.5° × 2.5° gridcells from 20° N to 60° N. An aggregate subset of 12 grid-cells was selected between 27.5° N and 42.5° N and within 5° longitude of the North American coast (Fig. 1) to isolate potential effects of eastern-seaboard pollution on oceanic precipitation. A fourth-root transformation of that data set produced the required gaussian distribution although our choice of transformation did not significantly affect our fundamental results. Aggregation removed potential spatial dependency problems<sup>21</sup>, and no difference in variance as a function of day was apparent.

For the hemispheric oceanic data set averaged over the 17-year time period, any given day comprises ~14.3% (one-seventh) of the weekly total. However, the aggregated near-coastal grid-cells show a distinctive intra-weekly variability. The greatest precipitation for these near-coastal grid-cells occur on Saturdays (658 mm yr<sup>-1</sup>, 16.0%) while Mondays receive only 538 mm yr<sup>-1</sup> (13.1%) (Fig. 2b). Consequently, Saturday precipitation averages 22% higher than Monday precipitation.

Statistical tests confirm these findings. First, a one-tailed *t*-test indicates that the difference in means between Monday and Saturday is significant. Second, when the satellite rainfall estimates are



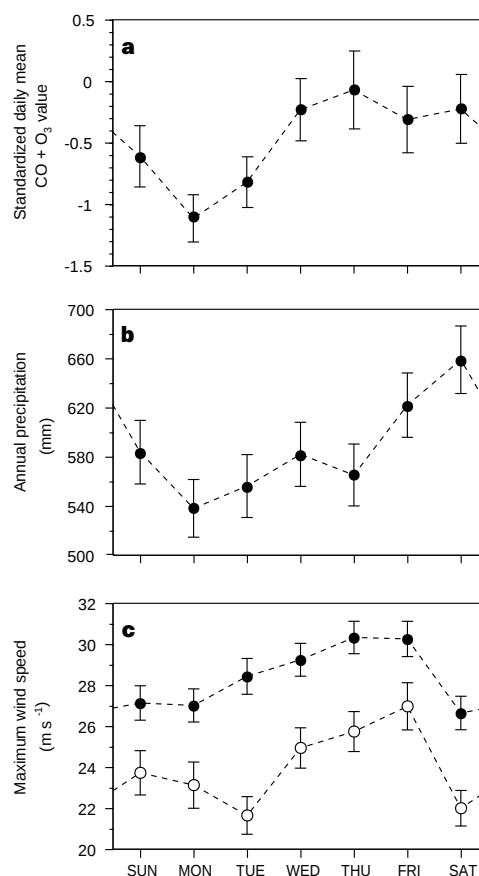
**Figure 1** Location map. In the main figure, an asterisk shows the position of Sable island, monitoring site for CO and O<sub>3</sub> data; also shown are the 12 oceanic satellite-derived rainfall grid-cells (heavy lines; within 5° of the North American coast and between 27.5° N and 42.5° N) and the 'coastal' region of tropical cyclone observations (hatching; north of 25° N, west of 75° W, east of 82° W). Inset shows the location of the study area.

categorized into 847 weekly periods, a 7-day sine wave (first harmonic) explains an average 40% of the variance in each weekly period. Third, spectral analysis confirms the existence of a seven-day periodicity.

Fourth, the weekly cycle is also apparent downwind (east) of these grid-cells. A set of grid-cells for the mid-Atlantic constructed as a 30° longitude offset from the near-coastal grid-cells shows a weekly cycle in total precipitation which is confirmed by spectral analysis. Fifth, harmonic analysis of the average value by day of the week for each 2.5° × 2.5° grid-cell from 20° N to 60° N in the Atlantic reveals a distinct phase shift. Maximum precipitation occurs near the weekend for coastal grid-cells, and progresses to midweek for the mid-Atlantic. Mid-Atlantic grid-cells have precipitation peaks occurring on Tuesdays (396.7 mm yr<sup>-1</sup>; 14.9%) and precipitation minima occurring on Saturdays (361.9 mm yr<sup>-1</sup>; 13.6%). This pattern is consistent with eastward transport of pollution plumes<sup>8,11</sup>.

A second independent test involves tropical cyclone analysis. Tropical cyclone data are derived from Atlantic observations (geographical positions, maximum wind strength, and minimum central pressure at six-hour intervals; 1886 to 1996) collected by the National Hurricane Center<sup>22</sup>. All observations are categorized by day of the week. Because of more limited hurricane data before 1946, only data from 1946 to 1996 are used.

An overestimation bias in reported winds exists between 1944 and 1969<sup>23</sup>. Observations for that period are stronger by 4.9 m s<sup>-1</sup> than observations between 1970 and 1996. Although the segregation by day mitigates the overestimation, two separate analyses were



**Figure 2** Mean values of three variables by day of the week; error bars show standard error of the mean ( $\pm 1\sigma$ ). **a**, Mean values for the summed composite of the standardized species (CO and O<sub>3</sub>) for Sable island, 1991–95 (the mean of the composite is nonzero owing to the exclusion of missing data pairs); **b**, total annual precipitation for grid-cells identified in Fig. 1; and **c**, tropical cyclonic mean maximum wind speed for the 'long' (filled circles) data set (1946–96) and 'short' (open circles) data set (1970–96).



conducted to fully compensate for this bias. The first analysis used the entire database ('long') while the second used only the unbiased (1970–96) observations ('short'). The total database showed no significant differences by day of the week in observation frequency, maximum surface wind speeds, or minimum central pressure. All tropical cyclone data sets met normality criteria.

Two regions are defined; 'coastal' incorporates all tropical cyclone observations north of 25° N, west of 70° W but east of 82° W (Fig. 1) and 'non-coastal' includes all tropical cyclone observations outside the 'coastal' parameters. As with the total data set, the 'non-coastal' data set shows no significant variation in strength or frequency by day of the week. For the 'non-coastal' data set, neither minimum central pressure nor maximum surface wind show weekly variability. Additionally, no differences exist in observation frequency by day of the week for the 'non-coastal' data set.

Conversely, the 'coastal' tropical cyclone data set shows an explicit weekly cycle (Fig. 2c). Both 'long' and 'short' 'coastal' weekend observations significantly vary from mid-week observations. For the 'long' data set, maximum surface winds average 5.0 m s<sup>-1</sup> slower for observations made on Saturdays than those for Fridays (Fig. 2c). The unbiased 'short' data set has Saturday wind observations averaging 3.6 m s<sup>-1</sup> slower than those on Friday; these differences are highly significant.

Significant differences are also apparent in minimum central pressures (not shown). The 'short' data set's Saturday observations average 6.3 hPa higher in central minimum pressure than for Thursday. Because of the discontinuous nature of the tropical cyclone data, spectral analysis is not applicable. However, when the maximum wind data are categorized into 396 separate weekly periods, a 7-day sine wave (first harmonic) explains an average 55% of the variance in each weekly period.

Because our goal is to emphasize the long-term existence, not direct comparison, of weekly cycles in three independent variables, caution is urged in applying day-to-day correspondences between the three weekly cycles. However, proving the existence of equivalent cycles is a time-honoured scientific methodology; for example, research linking cycles in palaeoclimate reconstructions with equivalent astronomical cycles<sup>24</sup>. Cycle identification is particularly important if a physical rationale can be advanced to explain cyclic similarities.

Theories proposed in hurricane modification studies<sup>25</sup> and urban climate modification studies<sup>3,26–28</sup> may explain the similar cycles in climate and pollution. Heavy aerosol loading can produce thermal responses in storm intensity and precipitation development<sup>2,25,29,30</sup> such that modification of the boundary layer's thermal structure leads to increased turbulence and enhanced precipitation<sup>29,30</sup>, and that solar radiation absorption by massive aerosol loading around a hurricane cloud cluster stimulates cumulus convection<sup>25</sup>. We propose that extensive regional pollution advection into the Atlantic produces climate modification by these mechanisms. As our findings are statistical in nature, we can not confirm such a cause-and-effect hypothesis, but we propose this hypothesis to stimulate research on potential causative mechanisms operating between pollution and regional climate. □

Received 30 December 1996; accepted 20 May 1998.

- Changnon, S. A. Jr. METROMEX: A review and summary. *Meteorol. Monogr.* **40**, 1–181 (1981).
- Dettwiller, J. W. & Changnon, S. A. Jr. Possible urban effects on maximum daily rainfall at Paris, St. Louis and Chicago. *J. Appl. Meteorol.* **15**, 517–519 (1976).
- Ackerman, B. S. *et al.* Summary of METROMEX Vol. 2, Causes of Precipitation Anomalies (Illinois State Water Survey, Urbana, IL, 1978).
- Houghton, J. T. *et al.* (eds) *Climate Change 1995: The Science of Climate Change* (Cambridge Univ. Press, 1996).
- Tett, S. F. B., Mitchell, J. F. B., Parker, D. E. & Allen, M. R. Human influence on the atmosphere vertical temperature structure: Detection and observations. *Science* **274**, 1170–1173 (1996).
- Gordon, A. H. Weekdays warmer than weekends? *Nature* **367**, 325–326 (1994).
- Graedel, T. E., Farrow, L. A. & Weber, T. A. Photochemistry of the 'Sunday Effect'. *Environ. Sci. Technol.* **11**, 690–694 (1977).
- Parrish, D. D. *et al.* Export of North American ozone pollution to the North Atlantic Ocean. *Science* **259**, 1436–1439 (1993).
- Pueschel, R. J., Boatman, J. F. & Artz, R. S. Aerosols over the western Atlantic: Scale heights, concentrations, and fluxes. *Atmos. Environ.* **22**, 2371–2380 (1988).
- Anderson, B. E. *et al.* The impact of U.S. continental outflow on ozone and aerosol distributions over

the Western Atlantic. *J. Geophys. Res.* **98**(D12), 23477–23489 (1993).

- Dickerson, R. R., Doddridge, B. G. & Kelley, P. Large-scale pollution of the atmosphere over remote Atlantic Ocean: Evidence from Bermuda. *J. Geophys. Res.* **100**, 8945–8952 (1995).
- Cleveland, W. S., Graedel, T. E., Kleiner, B. & Warner, J. L. Sunday and weekday variations in photochemical air pollutants in New Jersey and New York. *Science* **186**, 1037–1038 (1974).
- Altshuler, S. L., Arcado, T. D. & Lawson, D. R. Weekday vs. weekend ambient ozone concentration: Discussion and hypotheses with focus on Northern California. *Journal of Air and Waste Management Association* **45**, 967–972 (1995).
- Saeger, M. *et al.* *The 1985 NAPAP Emissions Inventory (Version 2): Development of the Annual Data and Modelers' Tapes* (US Environmental Protection Agency, Washington DC, 1989).
- Fishman, J. & Seiler, W. Correlative nature of ozone and carbon monoxide in the troposphere. Implications for the tropospheric ozone budget. *J. Geophys. Res.* **88**, 3662 (1983).
- Keeping, E. S. *Introduction to Statistical Inference* (Van Nostrand, Princeton, NJ, 1962).
- Siegel, S. *Non-Parametric Statistics* (McGraw-Hill, New York, 1956).
- Oltmans, S. J. & Levy, I. H. Seasonal cycle of surface ozone over the western North Atlantic. *Nature* **358**, 392–394 (1992).
- Ashworth, J. R. The influence of smoke and hot gases from factory chimneys on rainfall. *Q. J. R. Meteorol. Soc.* **55**, 841–850 (1929).
- Spencer, R. W. Global oceanic precipitation from the MSU during 1979–1991 and comparisons with other climatologies. *J. Clim.* **6**, 1301–1326 (1993).
- Livezey, R. E. & Chen, W. Y. Statistical field significance and its determination by Monte Carlo techniques. *Mon. Weath. Rev.* **111**, 46–59 (1983).
- Jarvinen, B. R., Neumann, C. J. & Davis, M. A. S. *A Tropical Cyclone Data Tape for the North Atlantic Basin, 1886–1983: Contents, Limitations, and Uses* (US Dept of Commerce, NOAA, Weather Service, Natl Hurricane Center, Coral Gables, Florida, 1984).
- Landsea, C. W. A climatology of intense (or major) Atlantic hurricanes. *Mon. Weath. Rev.* **121**, 1703–1713 (1993).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the earth's orbit: Pacemaker to the ice ages. *Science* **194**, 1121–1132 (1976).
- Gray, W. M., Frank, W. M., Corrin, M. L. & Stokes, C. A. Weather modification by carbon dust absorption of solar energy. *J. Appl. Meteorol.* **15**, 355–386 (1976).
- Gallo, K. P., Tarpley, J. D., McNab, A. L. & Karl, T. R. Assessment of urban heat islands: A satellite perspective. *Atmos. Res.* **37**, 37–43 (1995).
- Oke, T. R., Johnson, G. T., Steyn, D. G. & Watson, I. D. Simulation of surface urban heat islands under 'ideal' conditions at night, Part 2: diagnosis of causation. *Bound. Lay. Meteorol.* **36**, 339–358 (1991).
- Henry, J. A., Dicks, S. E., Wetterquist, O. F. & Roguski, S. J. Comparison of satellite, ground-based, and modeling techniques for analyzing the urban heat island. *Photogr. Eng. Remote Sensing* **55**, 69–76 (1989).
- Atkinson, B. W. The effect of an urban area on the precipitation from a moving thunderstorm. *J. Appl. Meteorol.* **10**, 47–55 (1971).
- Changnon, S. A. Jr. Recent studies of urban effects on precipitation in the United States. *Bull. Am. Meteorol. Soc.* **50**, 411–421 (1969).

**Acknowledgements.** We thank D. Parrish for the Canadian pollution data, S. J. Oltmans for the Bermuda ozone data, C. Landsea for the tropical cyclone database, and R. Spencer for the MSU daily precipitation data set.

Correspondence and requests for materials should be addressed to R.S.C. (e-mail: cerveny@asu.edu).

## Stable phytoplankton community structure in the Arabian Sea over the past 200,000 years

C. J. Schubert\*†, J. Villanueva\*†, S. E. Calvert\*, G. L. Cowie\*†, U. von Rad‡, H. Schulz‡, U. Berner‡ & H. Erlenkeuser§

\* Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

‡ Bundesanstalt für Geowissenschaften und Rohstoffe, D-30631 Hannover, Germany

§ Leibniz-Labor für Altersbestimmung und Isotopenforschung, Universität Kiel, D-24118 Kiel, Germany

Glacial to interglacial climate changes have been related to organic carbon cycling in oceanic surface waters<sup>1</sup>, and this possible link has led to the development of sedimentary tracers of past marine biological production. For example, sediment records of organic carbon<sup>2</sup>, opal<sup>3</sup> and biogenic barium<sup>4</sup> have been used to reconstruct past variations in production in different oceanic regimes, but these tracers cannot be used to discriminate between the relative contributions of different phytoplankton groups. Such a discrimination would provide greater insight into

† Present addresses: Max Planck Institute for Marine Microbiology, Celsiusstrasse 1, 28359 Bremen, Germany (C.J.S.); Département de Géologie et Océanographie, UMR CNRS 5805, Université de Bordeaux I, Avenue des Facultés, 33405 Talence Cedex, France (J.V.); Geology and Geophysics Department, University of Edinburgh, Edinburgh EH9 3JW, UK (G.L.C.).

the operation of the biological 'pump' transporting material down out of surface waters, and into the possible influence of the structure of oceanic food chains on carbon fluxes. Several organic biomarker compounds have now been established for tracing the contribution of different planktonic groups to organic carbon in sediments<sup>5–7</sup>. Here we show that four such biomarkers—dinosterol, alkenones, brassicasterol and chlorins, which represent dinoflagellates, prymnesiophytes, diatoms and chlorophyll-producers, respectively—have concordant concentration maxima that coincide with organic carbon maxima over the past 200,000 years in a sediment core from the northeastern Arabian Sea. Not only do these organic tracers track changes in ocean production in this region, but the similar distributions of dinosterol and brassicasterol indicate that the relative contributions of the dominant members of the phytoplankton community (diatoms and dinoflagellates) to production were roughly uniform on time-scales greater than 3,000–4,000 years over the past 200,000 years.

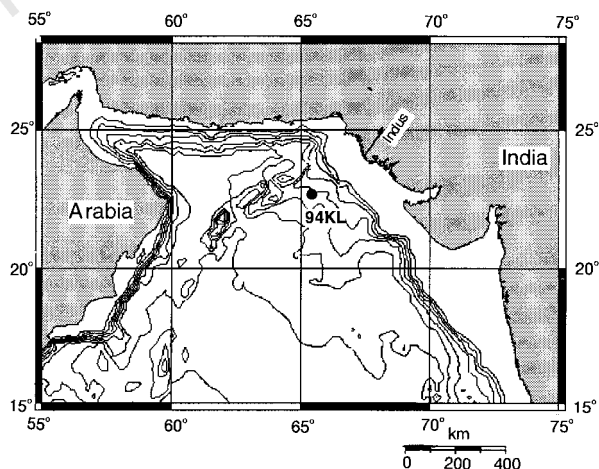
Biological production in the Arabian Sea is controlled by monsoon-driven upwelling<sup>8,9</sup>. Differential heating in the Indian Ocean and the Asian plateau leads to annual changes in atmospheric pressure over the sea and the land<sup>10</sup>. In summer, winds blow strongly from the ocean towards the Himalayas (the southwest monsoon), while in winter the pressure gradient (land–ocean) is reversed and winds blow from the continent to the sea (the northeast monsoon). The intensity of the monsoon system has varied strongly in the past on a precessional (23-kyr) periodicity<sup>11,12</sup>, and this has been well demonstrated from sediment records using production as well as lithogenic tracers<sup>12–14</sup>. Whereas the upwelling regime of the western Arabian Sea off Oman has been intensively investigated, only sparse data exist from the northeastern part of the Arabian Sea. In this area, high biological production occurs after the upwelling season and during much of the winter (northeast monsoon). Sea surface cooling during this period drives convection that leads to the injection of nutrients into the surface waters; such injection is the dominant control of primary production in winter<sup>15</sup>.

Core 94KL (22° 29.31' N, 65° 38.99' E) was taken from the continental margin off Pakistan in 2,109 m water depth<sup>16</sup> (Fig. 1). Age control over the past 200,000 years is provided by the oxygen isotope record of the planktonic foraminifera *Globigerinoides ruber* (Fig. 2a). The core location lies outside the influence of the Indus fan, thereby avoiding the effects of significant river-transported sediment, as shown by the uninterrupted hemipelagic sedimentation record<sup>16</sup>. Today the continental margin of the northeastern Arabian Sea is characterized by a distinct oxygen-minimum zone

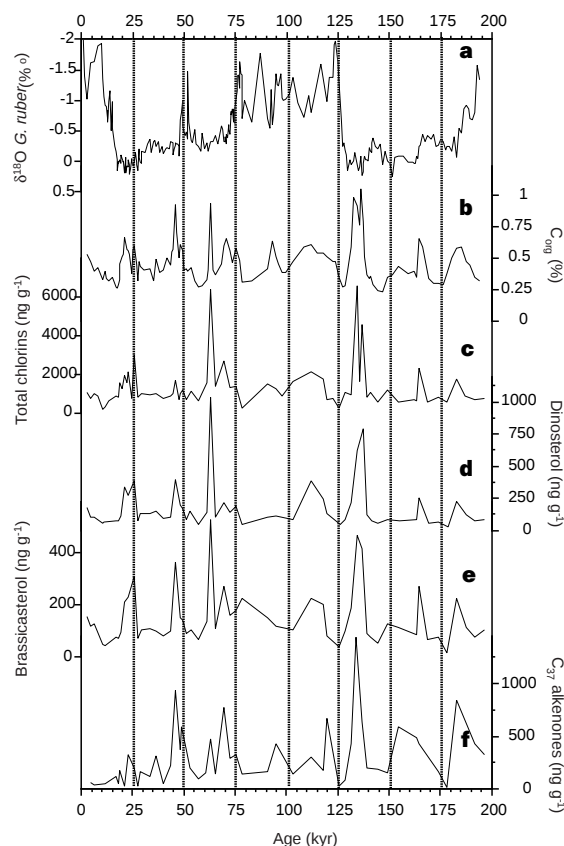
which extends from ~200 to 1,000 m water depth<sup>17</sup>. The highly bioturbated sediment sequence in core 94KL indicates that bottom waters were well oxygenated throughout the entire core record.

Organic carbon ( $C_{org}$ ) contents in the core range from 0.23 to 1.05 wt%, with maxima every 18–28 kyr (Fig. 2b). Assuming our stratigraphic age model has an uncertainty of  $\pm 5$  kyr (ref. 18), these changes may well reflect a 23-kyr cycle. The organic material is primarily of marine origin throughout the core, as shown by  $\delta^{13}C_{org}$  values that range from –19.4‰ to –21.5‰ versus VPDB, a  $\delta^{13}C$  with concentrations of uniquely terrestrial lignin phenols (0.05–0.4 ng per 100 mg  $C_{org}$ ). Also shown in Fig. 2 are the concentration profiles of four different biomarkers and biomarker classes. Total sedimentary chlorins, the transformation products of chlorophyll, have recently been used as a taxonomically independent palaeoproduction indicator in upwelling areas off Africa<sup>19</sup>. That is to say, their abundances in a sediment may be used as an index of varying export production. Dinosterol is a well established biomarker of dinoflagellates<sup>6</sup> whose sedimentary concentration has been correlated with changes in marine production rate<sup>20</sup>. Brassicasterol is a biomarker indicating production changes related to diatom abundances<sup>21</sup>. Alkenones are restricted to prymnesiophycean algae<sup>22</sup> and their abundances have also been used as production indicators<sup>7,14</sup>.

In core 94KL, total chlorin concentrations vary from 199 to 6,515 ng g<sup>–1</sup>, with maxima at almost all  $C_{org}$  maxima (Fig. 2c), the principal exceptions being at 45 and 91 kyr. The dinosterol and brassicasterol concentrations are also well correlated with  $C_{org}$



**Figure 1** Location of core 94KL, off Pakistan in the northeastern Arabian Sea. The core is located below the intense oxygen-minimum zone on this margin. The organic geochemical subsamples were deep-frozen immediately upon recovery to prevent decomposition of organic compounds.

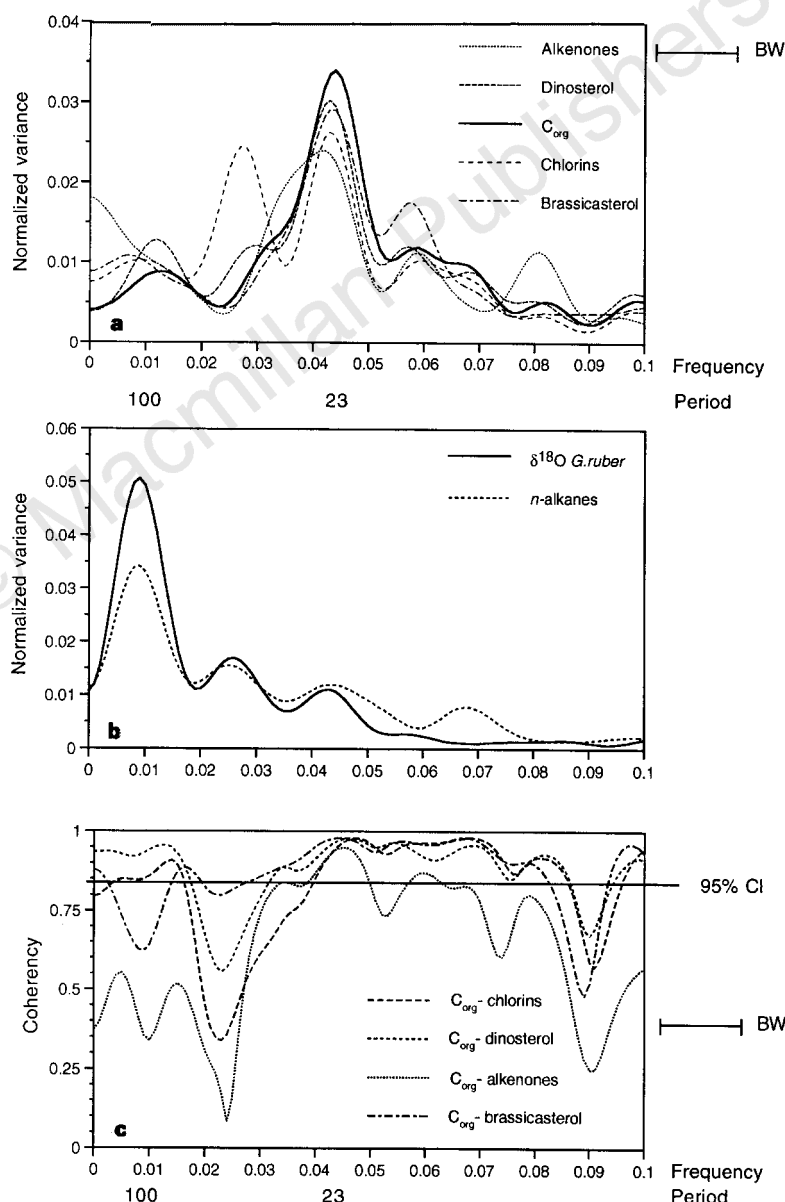


**Figure 2** Data from core 94KL, plotted against age. **a**, Oxygen isotope ratio of *Globigerinoides ruber* (‰ versus VPDB, size fraction 315–400  $\mu$ m); **b**, organic carbon (per cent dry weight of sediment); **c**, total chlorin concentration (ng per g dry weight of sediment); **d**, dinosterol concentration; **e**, brassicasterol concentration; **f**,  $C_{37}$  alkenones concentration. Organic carbon, dinosterol, brassicasterol, alkenone and chlorin concentrations were determined at 5 to 20 cm intervals corresponding to a resolution of 900–3,600 yr. Experimental details are given in Methods.

(Fig. 2d and e). Although dinosterol and brassicasterol concentration maxima are absent at 91 kyr, the excellent correlations between concentrations of total chlorins and dinosterol ( $r = 0.91$ ) and between the concentration of total chlorins and brassicasterol ( $r = 0.90$ ) suggest nevertheless that dinoflagellates and diatoms have been the main contributors to chlorophyll production in this area. This is supported by the highest peaks in the dinosterol and brassicasterol profiles around 64 kyr which coincide with significant concentration peaks in the profiles of  $C_{org}$  and total chlorins. Alkenones range in concentration from 16 to  $1,433 \text{ ng g}^{-1}$  and show a somewhat weaker correlation with the  $C_{org}$  profile ( $r = 0.68$ ; Fig. 2f), but again the principal concentration maxima occur at the  $C_{org}$  maxima. These correlations are higher during the past 150 kyr than during the period 150–200 kyr, implying that the planktonic community varied more strongly before 150 kyr. We note that the double peak in  $C_{org}$  and total chlorin concentrations at

~135 kyr can be resolved into a high dinosterol contribution (first peak) and high alkenone and brassicasterol contributions (second peak). This demonstrates the considerable potential of the multiple biomarker method presented here, as it is able to distinguish between different total organic carbon contributors.

Spectral analysis<sup>23,24</sup> of the  $\delta^{18}\text{O}$  (*G. ruber*),  $C_{org}$ , dinosterol, brassicasterol, alkenone, long-chain *n*-alkanes, and chlorin profiles for core 94KL are shown in Fig. 3a and b. Spectral peaks for the marine organic parameters range from 0.041 to 0.044, with a mean value of 0.043 (23-kyr periodicity), consistent, given our chronological uncertainty, with mean precession (Fig. 3a). This is consistent with earlier investigations that demonstrated a close coupling of production variations to precessionally driven variations in monsoon intensity<sup>12–14</sup>. Geochemical profiles of terrestrial organic matter supply, such as long-chain *n*-alkanes and *n*-alcohols (C.J.S. and J.V., unpublished data) do not coincide with the marine proxies



**Figure 3** Time series analysis. **a**, Spectral analysis<sup>23,24</sup> of the  $C_{org}$ , total chlorins, dinosterol, brassicasterol and alkenone profiles of core 94KL. The marine biomarkers show a range of peak values from 0.041 to 0.044 with a mean value of 0.043 (that is, 23-kyr cyclicality) consistent, given our chronological uncertainty, with mean precession. **b**, Spectral analysis of  $\delta^{18}\text{O}$  (*Globigerinoides ruber*) and long-chain *n*-alkane profiles, showing a dominant 100-kyr cyclicality with decreasing

power at 41 kyr and 23 kyr. **c**, Cross-spectral analysis of  $C_{org}$  and biomarkers, showing very high values at the 23-kyr band, indicating a strong coherence between the overall organic matter input and the main planktonic producers (95% confidence level is 0.810). 'BW' indicates band-width, 'CI' indicates confidence-level; 'Frequency' is given in units of cycle kyr<sup>-1</sup>, and 'Period' in kyr.



used in this study but vary on a glacial/interglacial cycle. This is shown in Fig. 3b where the  $\delta^{18}\text{O}$  and long-chain *n*-alkane records show a dominant 100-kyr periodicity with decreasing power at 41 kyr and 23 kyr, corresponding with the ice volume record<sup>25</sup>. The very high spectral coherencies between biomarkers and  $C_{\text{org}}$  (0.934 for alkenones and pigments, 0.960 for dinosterol and 0.977 for brassicasterol, with a 95% confidence level of 0.810) strongly support the close correlation described above between the overall marine production ( $C_{\text{org}}$ ) and specific plankton groups (Fig. 3c). No significant shift is apparent in the phase relation between the parameters; all phase angles at 23 kyr are smaller than  $16^\circ$  (that is, 1.1 kyr), suggesting a uniform development of the planktonic community over time.

The observed changes in biomarker abundances in core 94KL are unlikely to be due to temporal variations in preservation in view of the location of the core (well below the depth of the oxygen minimum) and the fact that it is bioturbated throughout. On the other hand, we are unable to rule out differences in preservation between the different classes of organic compounds with the data available. Nevertheless, the significant difference in the records of long-chain *n*-alkanes and *n*-alcohols on the one hand, and alkenones, dinosterol and brassicasterol on the other, suggests that the variability of the organic geochemical signals is not significantly affected by preservation artefacts.

Primary production in the central and northern Arabian Sea at present is predominantly due to dinoflagellates, diatoms and coccolithophorids<sup>26,27</sup>. The almost identical profiles of the abundances of dinosterol and brassicasterol ( $r = 0.92$ ) in core 94KL suggest that the relative contributions of these algal groups to primary production has not changed over the past 200,000 years, although overall palaeoproduction has changed dramatically (estimated palaeoproduction values<sup>28</sup> (derived from  $C_{\text{org}}$ ) range from 70 to 200  $\text{g C m}^{-2} \text{yr}^{-1}$ ). Moreover, the strongly coherent chlorin profile supports the suggestion<sup>19</sup> that this proxy is a useful tool for monitoring total primary production variations. We therefore conclude that although the rate of marine production in the eastern Arabian Sea has changed markedly over the last six climate cycles (assuming variations in preservation to be negligible), this was not due to a change in the structure of the planktonic community; increased upwelling and nutrient supply due to increased wind speeds evidently affected a wide range of planktonic organisms, including the diatoms, the group normally assumed to respond most dramatically to varying nutrient supply<sup>29</sup>. □

## Methods

Organic carbon was determined on carbonate-free sub-samples using a Carlo Erba NA-1500 CHN analyser. For the determination of alkenones, brassicasterol, dinosterol and *n*-alkanes, freeze-dried sediment ( $\sim 2 \text{ g}$ ) was extracted by successive sonication and centrifugation in methanol, methanol:methylene chloride (1:1), and methylene chloride. The extracts were saponified (6% KOH), further extracted with hexane, and derivatized with *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA) before injection on a HP1 chromatographic column (50 m long, 0.32 mm i.d., 0.17  $\mu\text{m}$ ). For chlorin measurements, freeze-dried sediment sub-samples ( $\sim 0.6 \text{ g}$ ) were extracted by threefold sonication and centrifugation in acetone. The samples were cooled with ice under low light conditions during extraction to prevent decomposition of chlorin. Sediment extracts were measured fluorimetrically immediately after extraction using a Turner fluorometer. Lignin-derived phenols were determined by gas chromatography-flame ionization detection of the trimethylsilyl derivatives following CuO oxidation of bulk samples under alkaline conditions<sup>30</sup>. The precision of the organic geochemical methods used is  $\pm 10$ –15%.

Received 27 October 1997; accepted 12 May 1998.

1. Sarnthein, M., Winn, K., Duplessy, J. C. & Fontugne, M. R. Global variations and surface ocean productivity in low and mid latitudes: Influence on  $\text{CO}_2$  reservoirs of deep ocean and atmosphere during the last 21,000 years. *Paleoceanography* **3**, 361–399 (1988).
2. Müller, P. J. & Suess, E. Productivity, sedimentation rate, and sedimentary organic matter in the oceans. *Deep-Sea Res.* **26**, 1347–1362 (1979).

3. DeMaster, D. J. The supply and accumulation of silica in the marine environment. *Geochim. Cosmochim. Acta* **45**, 1715–1732 (1981).
4. Dymond, J., Suess, E. & Lyle, M. Barium in deep-sea sediment: A geochemical proxy for paleoproductivity. *Paleoceanography* **7**, 163–181 (1992).
5. Brassell, S. C. & Eglinton, G. in *Molecular Geochemical Indicators in Sediments in Organic Marine Geochemistry* (ed. Sohn, M. L.) 10–32 (Am. Chem. Soc., Washington DC, 1986).
6. Boon, J. J., Rijpstra, W. I. C., Lange, F. & de Leeuw, J. W. Black Sea sterol—a molecular fossil for dinoflagellate blooms. *Nature* **277**, 125–127 (1979).
7. Brassell, S. C., Eglinton, G., Marlowe, I. T., Pflaumann, U. & Sarnthein, M. Molecular stratigraphy: a new tool for climatic assessment. *Nature* **320**, 129–133 (1986).
8. Nair, R. R. et al. Increased particle flux to the deep ocean related to monsoons. *Nature* **338**, 749–751 (1989).
9. Brock, J. C. & McClain, C. R. Interannual variability in phytoplankton blooms observed in the northwestern Arabian Sea during the southwest monsoon. *J. Geophys. Res.* **97**, 733–750 (1992).
10. Kutzbach, J. E. Monsoon climate of the Early Holocene: climate experiment with the Earth's orbital parameters for 9,000 years ago. *Science* **214**, 59–61 (1981).
11. Prell, W. L. & Kutzbach, J. E. Sensitivity of the Indian monsoon to forcing parameters and implications for its evolution. *Nature* **360**, 647–651 (1992).
12. Clemens, S., Prell, W., Murray, D., Shimmield, G. & Weedon, G. Forcing mechanisms of the Indian Ocean monsoon. *Nature* **353**, 720–725 (1991).
13. Shimmield, G. B. & Mowbray, S. T. The inorganic geochemical record of the northwestern Arabian Sea: a history of productivity variation over the last 400 k.y. from sites 722 and 724. *Proc. ODP Sci. Res.* **117**, 409–419 (1991).
14. Rostek, F. et al. in *Global Precipitations and Climate Change* (eds Desbois, M. & Désalmand, F.) 27–51 (Springer, Berlin, 1994).
15. Madhupratap, M. et al. Mechanism of the biological response to winter cooling in the northeastern Arabian Sea. *Nature* **384**, 549–552 (1996).
16. von Rad, U. et al. Influence of the Oxygen Minimum Zone on the Sedimentation at the Upper Continental Slope off Pakistan, "Pakomir" (Sonne Cruise SO90 Cruise Rep., 1993).
17. Wyrtki, K. *Oceanographic Atlas of the International Indian Ocean Expedition* (NSF, Washington DC, 1971).
18. Martinson, D. G. et al. Age dating and the orbital theory of the ice ages: Development of a high-resolution 0 to 300,000 year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
19. Harris, P. G. et al. Chlorin accumulation rate as a proxy for Quaternary marine primary productivity. *Nature* **383**, 63–65 (1996).
20. Prahl, F. G., Muehlhausen, L. A. & Lyle, M. An organic geochemical assessment of oceanographic conditions at MANOP site C over the past 26,000 years. *Paleoceanography* **4**, 495–510 (1989).
21. Volkman, J. K. A review of sterol markers for marine and terrigenous organic matter. *Org. Geochem.* **9**(2), 83–99 (1986).
22. Marlowe, I. T. et al. Long chain (*n*-C37–C39) alkenones in the Prymnesiophyceae. Distribution of alkenones and other lipids and their taxonomic significance. *Br. Phycol. J.* **19**, 203–216 (1984).
23. Jenkins, G. M. & Watts, D. G. *Spectral Analysis and its Applications* (Holden Day, Oakland, 1968).
24. Paillard, D. & Labeyrie, L. D. A "user-friendly" Macintosh software for rapid correlations of paleoclimatic signals and treatments (Centre des Faibles Radioactivités PNEDC/INSU, 1993).
25. Imbrie, J. et al. in *Milankowitch and Climate Pt 1* (ed. Berger, A. L.) 307–330 (1984).
26. Krey, J. in *The Biology of the Indian Ocean* (ed. Zeitzschel, B.) 115–126 (Springer, Berlin, 1973).
27. Pollehne, F., Klein, B. & Zeitzschel, B. Low light adaption and export production in the deep chlorophyll maximum layer in the northern Indian Ocean. *Deep-Sea Res.* **II** **40**, 737–752 (1993).
28. Stein, R. in *Biochemistry of Black Shales* (eds Degens, E. T. et al.) 55–70 (Mitt. Geol. Paläont. Inst., Univ. Hamburg, 1986).
29. Wangersky, P. J. The role of particulate matter in the productivity of surface waters. *Helgoländer Wiss. Meer.* **30**, 546–564 (1977).
30. Hedges, J. E. & Ertel, J. R. Characterization of lignin by gas capillary chromatography of cupric oxide oxidation products. *Anal. Chem.* **54**, 174–178 (1982).

**Acknowledgements.** We thank P. Parkinson (Civil and Mechanical Engineering, UBC) for providing access to the GC/MS system. C.J.S. was supported by a NATO fellowship provided by the German Academic Exchange Service (DAAD). J.V. was supported by the Spanish Ministry of Education. Laboratory support was provided by NSERC (S.E.C.).

Correspondence and requests for materials should be addressed to C.J.S. (cschuber@mpi-bremen.de).

## Melting of a subducting oceanic crust from U–Th disequilibria in austral Andean lavas

O. Sigmarsson\*†, H. Martin\* & J. Knowles\*

\* Centre National de la Recherche Scientifique (CNRS) and Université Blaise Pascal, 5 rue Kessler, 63038 Clermont-Ferrand, France

† Earth Sciences Department, University of California, Santa Cruz, California 95064, USA

Understanding crustal genesis at convergent plate boundaries is important for determining mass transfer between different geochemical reservoirs in the Earth's mantle, and for deciphering the long-term growth of the continental crust. Most arc magmas are thought to be generated from fluid-induced melting of the mantle wedge above slabs of subducting oceanic crust<sup>1</sup>. Such magmas frequently display <sup>238</sup>U enrichments or radioactive equilibrium<sup>2,3</sup>

between  $^{238}\text{U}$  and its radiogenic product  $^{230}\text{Th}$ . But where a young and hot oceanic crust is being subducted it may itself partially melt and produce calc-alkaline andesites and dacites, termed adakites<sup>4</sup>. Here we report a uniform excess of  $^{230}\text{Th}$  over  $^{238}\text{U}$ , but variable Th isotope ratios, in young adakites from the Andean austral volcanic zone south of the triple junction where the Chile ridge subducts beneath South America. We show that these results are compatible with the adakites having been formed by approximately 20% equilibrium melting due to amphibole decomposition in a heterogeneous<sup>5</sup> oceanic crust. Moreover, both the degree of melting of the oceanic crust and its thermal structure appear to be uniform under most of the Andean austral volcanic zone. Such partial melting of subducted oceanic slabs may have occurred throughout the Earth's history where young oceanic plates were subducted.

Adakites are characterized by high La/Yb and Sr/Y ratios but low Yb and Y contents which suggest residual garnet after partial melting, and they are compositionally similar to Archaean trondhjemite-tonalite-granodiorite (for example, refs 4 and 6). Due to the contrasting chemical properties of U and Th, radioactive disequilibrium between  $^{238}\text{U}$  and  $^{230}\text{Th}$  in arc lavas is a useful tracer of recent magma generation processes. Arc lavas frequently display  $^{238}\text{U}$ -enrichments due to recent slab-derived fluid addition to the mantle source<sup>4,7</sup>, whereas radioactive equilibrium between  $^{230}\text{Th}$  and  $^{238}\text{U}$  may reflect melting of metasomatic phases from former fluid infiltration to the mantle wedge above subducting slabs<sup>3</sup>. Here we present U–Th disequilibrium results for adakites, from the Andean austral volcanic zone (AVZ) in southern Chile, in order to test the slab melting model for their origin.

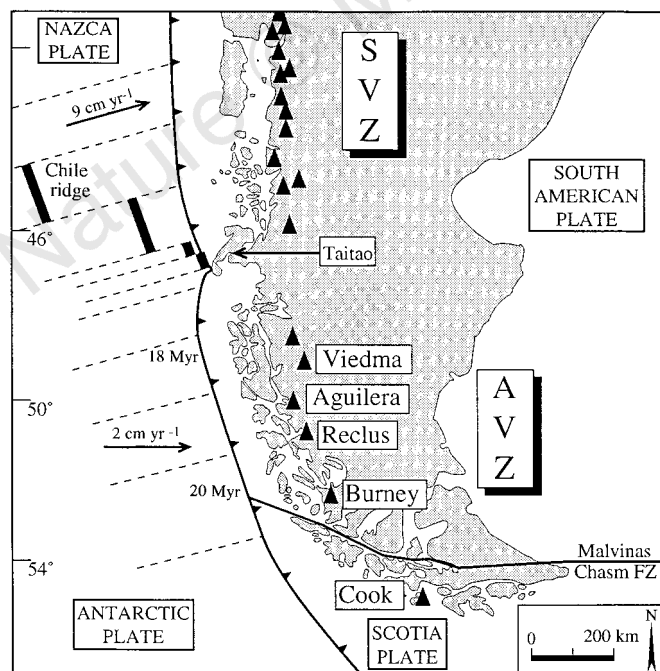
In the southern Andes, the Chile ridge is being subducted at the Chile ridge–trench triple junction close to the Taitao peninsula (Fig. 1). North of the triple junction, the Nazca plate subducts at the rate of  $9\text{ cm yr}^{-1}$  beneath the southern volcanic zone (SVZ) on the South American plate, whereas to the south the Antarctica plate is being subducted at a rate of only  $2\text{ cm yr}^{-1}$  (ref. 8). On the Taitao

peninsula an ophiolite complex<sup>9,10</sup> has been found (intruded by adakitic plutons<sup>11</sup>), implying that occasionally a part of the oceanic plate is obducted rather than subducted. The age of the Antarctica plate at the Chile trench where it subducts beneath the AVZ ( $49\text{--}55^\circ\text{S}$ ) is between 18 and 20 Myr (magnetic anomalies 5e to 6) and decreases down dip<sup>8</sup>. Little is known about the geological structure beneath the AVZ owing to the lack of seismicity, attributed to a relatively hot oceanic crust and a slow convergence rate<sup>12</sup>. Stern and Kilian<sup>13</sup> suggested that the crust is less than 35 km thick beneath the AVZ, consistent with their observation of only plagioclase and pyroxene granulites without garnet in xenoliths of lower-crustal origin which, therefore, cannot be the source for the adakites. A continuous suite of calc-alkaline basalts to rhyolites<sup>14</sup> is produced in the SVZ, whereas no basalts but only adakitic andesites and dacites are known from the AVZ<sup>13,15,16</sup>. To the extreme south of the AVZ, the Cook volcano is located on a different tectonic plate, the Scotia plate, and at a significantly greater distance from the Chile trench.

Our samples are Holocene in age; dacites from Viedma, Aguilera, Reclus and Burney volcanoes from the AVZ, and a high-MgO andesite from Cook Island. The samples consist of plagioclase, pyroxenes and amphibole phenocrysts, and groundmass containing fresh glass<sup>13</sup>. None of these minerals will significantly fractionate Th/U. The dacites show a restricted range in  $\text{SiO}_2$  contents from 64 to 66% (ref. 13), but their Th contents vary by more than an order of magnitude. The Th contents, and the  $(^{230}\text{Th}/^{232}\text{Th})$  and Th/U ratios, show a regular variation along the strike of the AVZ (Table 1). From north to south, the Th abundance decreases from 11 p.p.m. at Aguilera to less than 1 p.p.m. at Burney,  $(^{230}\text{Th}/^{232}\text{Th})$  increases from 0.66 in Viedma to 0.97 in Burney, and Th/U decreases from 5.00 to 3.37 (Fig. 2). Similar geographical variations have been observed for Sr, Nd, Pb and O isotope ratios<sup>13,15</sup> (for example,  $^{87}\text{Sr}/^{86}\text{Sr} = 0.7051\text{--}0.7028$ ). In contrast with the along-strike variability of isotope and trace-element ratios and Th contents, all the AVZ dacites have indistinguishable  $(^{238}\text{U}/^{230}\text{Th})$  of  $0.92 \pm 0.02$ . The Cook andesite has distinct U–Th systematics without any clear relationship to the other AVZ lavas.

The limited range in isotope and trace-element ratios in several samples from individual AVZ volcanoes<sup>13</sup> suggests that the large variability of  $(^{230}\text{Th}/^{232}\text{Th})$  and Th/U in the adakites (Fig. 2) is unlikely to result from intra-volcanic differentiation, but instead is probably inherited from the source region. In the light of the regular spatial variations of several isotope and trace-element ratios, crustal contamination appears unlikely. Moreover, such contamination would result in decreasing  $^{238}\text{U}\text{--}^{230}\text{Th}$  disequilibria (ref. 17), which is in contrast with the AVZ results. The difference from the SVZ, where relatively homogenous  $(^{230}\text{Th}/^{232}\text{Th})$  values were observed<sup>18</sup>, suggests a more variable source for the AVZ lavas. The source for the AVZ lavas could well be the subducted oceanic plate<sup>13,15,16</sup>, the composition of which can be deduced from the recent results<sup>5</sup> on the Chile ridge basalts. Indeed, these basalts have an uncommonly large range of Th/U (Fig. 2) which exceeds that of the AVZ magmas. Long-lived isotope ratios of the AVZ lavas<sup>13</sup> are also mostly in the range of the few analysed Chile ridge basalts<sup>5</sup>. Seawater alteration and sediment incorporation would lead to even larger isotope variability of the oceanic crust. Consequently, the Chile ridge basalts could be parental to AVZ magmas, and the regular geographical variations of the AVZ lava compositions suggest that the composition of the subducted plate varies considerably from north to south. The reason for the large heterogeneity and arc-like trace-element signature of the Chile ridge basalts is unclear<sup>5</sup>, but during occasional ridge obduction, and consequent formation of ophiolites, the convecting mantle wedge may have been diverted into the source region of these peculiar basalts.

Partial melting of a heterogeneous oceanic crust producing the AVZ adakites can be shown to be fully consistent with the U–Th disequilibrium results. Thermal models for a slow subduction of a young oceanic plate suggest that this melting could occur under



**Figure 1** Map of the southern Andes showing location of the volcanoes discussed in the text (modified from ref. 13). The austral volcanic zone (AVZ) and the southern volcanic zone (SVZ) are respectively to the south and the north of the Chile ridge–trench triple junction, which is currently located close to the Taitao peninsula. The subduction rates and the age of the subducting plates are from ref. 8.

**Table 1 Analytical results**

| Volcano         | U ( $\mu\text{g g}^{-1}$ ) | Th ( $\mu\text{g g}^{-1}$ ) | Th/U | ( $^{238}\text{U}/^{232}\text{Th}$ ) | ( $^{230}\text{Th}/^{232}\text{Th}$ ) | ( $^{238}\text{U}/^{230}\text{Th}$ ) | $^{87}\text{Sr}/^{86}\text{Sr}$ | SiO <sub>2</sub> (wt%) |
|-----------------|----------------------------|-----------------------------|------|--------------------------------------|---------------------------------------|--------------------------------------|---------------------------------|------------------------|
| Viedma          | 1.52                       | 7.61                        | 5.00 | 0.606                                |                                       |                                      | 0.70511 $\pm$ 2                 | 64.6                   |
| Viedma-d        | 1.52                       | 7.60                        | 5.00 | 0.607                                |                                       |                                      |                                 |                        |
| Aguilera        | 2.45                       | 11.0                        | 4.47 | 0.679                                | 0.732 $\pm$ 10                        | 0.93                                 | 0.70493 $\pm$ 2                 | 64.0                   |
| Reclus          | 0.759                      | 3.01                        | 3.96 | 0.766                                | 0.825 $\pm$ 15                        | 0.92                                 | 0.70504 $\pm$ 2                 | 65.8                   |
| Burney 131289-1 | 0.363                      | 1.23                        | 3.38 | 0.897                                | 0.972 $\pm$ 20                        | 0.93                                 | 0.70427 $\pm$ 2                 | 64.0                   |
| Burney-d        | 0.363                      | 1.22                        | 3.37 | 0.900                                |                                       |                                      |                                 |                        |
| Burney 131285-4 | 0.263                      | 0.950                       | 3.62 | 0.839                                | 0.912 $\pm$ 20                        | 0.90                                 | 0.70414 $\pm$ 2                 | 64.0                   |
| Burney 180290-2 | 0.304                      | 1.12                        | 3.67 | 0.827                                |                                       |                                      |                                 | 64.0                   |
| Cook            | 0.797                      | 3.48                        | 4.37 | 0.694                                | 0.883 $\pm$ 13                        | 0.79                                 | 0.70283 $\pm$ 2                 | 60.0                   |
| Cook-d          | 0.796                      | 3.47                        | 4.36 | 0.696                                |                                       |                                      |                                 |                        |

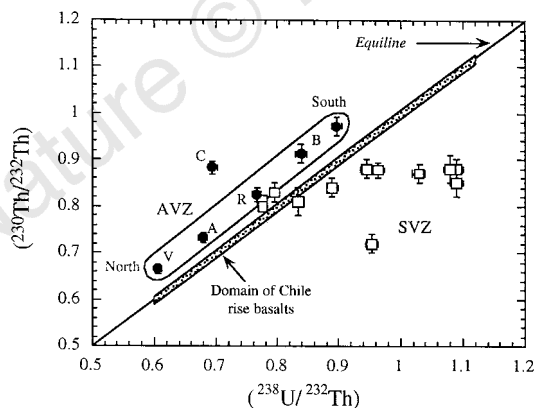
Duplicates are denoted by "-d". SiO<sub>2</sub> is from ref. 13. U, Th and  $^{87}\text{Sr}/^{86}\text{Sr}$  are from mass spectrometry and ( $^{230}\text{Th}/^{232}\text{Th}$ ) by  $\alpha$ -spectrometry. Parentheses around nuclides denote activity. Analytic errors are  $2\sigma_m$  precisions and are estimated as 0.5% for U and Th contents.  $^{87}\text{Sr}/^{86}\text{Sr}$  are normalized to  $^{86}\text{Sr}/^{86}\text{Sr} = 0.1194$  and  $^{87}\text{Sr}/^{86}\text{Sr} = 0.71022$  for the NSB-987 standard. All samples are <10,000 yr old. Possible post-eruptive decay during 10,000 yr would decrease the ( $^{230}\text{Th}/^{232}\text{Th}$ ) by  $\sim 1\%$  which is insignificant compared to the analytical errors.

amphibolite or eclogite conditions<sup>19</sup>. The solid residue should be principally composed of garnet (gt) and clinopyroxene (cpx) to be consistent with the distinctive high La/Yb in adakites<sup>4</sup>, with the composition of both liquids from amphibolite melting experiments<sup>20</sup>, and with SiO<sub>2</sub>-rich glass inclusions in peridotite nodules<sup>21</sup>. Moreover, abundant gt in eclogitic residue would cause Th enrichment in the melt because of the larger U partition coefficient for gt;  $D_U > D_{Th}$  (refs 22, 23). Adakitic melts have been experimentally generated at 3.2 GPa and 1,100 °C after  $\sim 20\%$  of amphibolite melting, with 42% gt, 37% cpx and 1% rutile in the residue<sup>20</sup>. Using these results and partition coefficients from refs 22 and 24, and assuming negligible  $D_{Th}$ ,  $D_{La}$  and  $D_{Yb}$  (ref. 25) and  $D_U$  of unity between rutile and melt (inferred from U = 11–390 p.p.m. and Th/U < 0.1 in rutile from eclogites<sup>26,27</sup>), we obtain 8% excess  $^{230}\text{Th}$  as observed in the AVZ lavas (that is, ( $^{238}\text{U}/^{230}\text{Th}$ ) = 0.92). The  $^{230}\text{Th}$  excess depends on degree of melting and amount of residual garnet, but is independent of source composition, in contrast to Th/U and La/Yb which depend on all three factors. The Chile rise basalts have La/Yb ranging from 0.9 to 3.68 (ref. 5), which upon  $\sim 20\%$  partial melting would yield La/Yb of 14 and 59 in the melts,

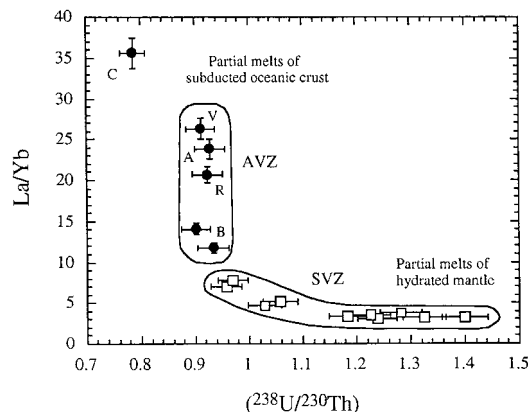
respectively. This range envelopes the measured range for all the AVZ lavas<sup>13</sup> (Fig. 3). Furthermore, a 20% melting of the Chile rise basalts would generate Sr/Y between 40 and 160 similar to those measured in the AVZ adakites<sup>13</sup>, but this ratio will be largely affected by plagioclase fractionation or accumulation at higher levels.

The subducted oceanic crust, regardless of its variable composition, was probably in secular equilibrium (that is, ( $^{238}\text{U}/^{230}\text{Th}$ ) = 1). In such case, the constancy of ( $^{238}\text{U}/^{230}\text{Th}$ ) in all AVZ dacites strongly suggests an origin from a uniform degree of slab melting, probably due to amphibole breakdown, leaving similar eclogitic residues for further subduction. This indicates an approximately uniform slab thermal structure under most of the AVZ, which is consistent with the oceanic crust being of a comparable age. However, a significantly lower degree of melting of a relatively dehydrated oceanic crust with, therefore, little modal amphibole is implied by the higher  $^{230}\text{Th}$  excess and La/Yb in the Cook andesite. Eventual assimilation of peridotite when the slab melt moves up through hotter mantle wedge<sup>28</sup> would not affect the above considerations because of low incompatible-element concentrations in the main mantle minerals.

The  $^{238}\text{U}$ – $^{230}\text{Th}$  disequilibria data thus show a regular but



**Figure 2** The Th–U isochron diagram showing the contrasting results for AVZ and SVZ. Aguilera, Burney, Cook, Reclus and Viedma volcanoes are denoted respectively as A, B, C, R and V. The SVZ data are from ref. 18 and O.S. *et al.*, unpublished results. The range of Th/U measured in basalts from the Chile ridge<sup>5</sup> is shown on the equiline ( $\text{Th}/\text{U} = 3.034/(\text{^{238}U}/\text{^{232}Th})$ ). This large range is consistent with the AVZ lavas being derived from partial melting of the subducting oceanic crust produced at the Chile rise.



**Figure 3** Variations of La/Yb with ( $^{238}\text{U}/^{230}\text{Th}$ ). Abbreviations are the same as in Figs 1 and 2. The La/Yb from the AVZ are from ref. 13, and the SVZ data are from ref. 18 and O.S. *et al.*, unpublished results. The U-excesses in SVZ lavas have been ascribed to fluid-induced melting of the mantle wedge<sup>18</sup>, which had little effect on La/Yb. In marked contrast, the constant  $^{230}\text{Th}$ -excess in the AVZ, excluding Cook, probably reflects uniform partial melting of the young Antarctica plate. Shallow dehydration of the hot plate during subduction would explain the absence of fluid signature in the AVZ lavas<sup>29</sup>. The compositions of the Chile rise basalts<sup>5</sup> strongly suggest that the subducting oceanic crust is highly heterogeneous in composition, which is supported by the large variations of La/Yb in AVZ lavas<sup>13</sup>. This heterogeneity has no effect on ( $^{238}\text{U}/^{230}\text{Th}$ ) which would be equal to unity in the slab despite variable Th/U and ( $^{230}\text{Th}/^{232}\text{Th}$ ). The partition coefficients used for garnet, clinopyroxene and rutile in the slab-melting calculation are respectively:  $D_{Th} = 0.0015, 0.0013, 0$ ;  $D_U = 0.01, 0.013, 1$ ;  $D_{La} = 0.0016, 0.05, 0$ ; and  $D_{Yb} = 6.6, 0.432, 0$  (refs 22–25). Other melting parameters are given in the text.



contrasting geochemical pattern in lavas from the SVZ and the AVZ, reflecting clear differences in petrogenesis to the north and south of the Chile triple junction. In the SVZ, variable enrichment of  $^{238}\text{U}$ , but uniform La/Yb, are attributed to slab-derived fluids and mantle melting (Fig. 3), whereas in the AVZ, the uniform  $^{238}\text{U}$  deficiency but high and variable La/Yb could result from  $\sim 20\%$  partial melting of the young and heterogeneous, subducted, Antarctica plate.

The compositional similarities of adakites and Archaean trondhjemites suggest that the early continental crust is likely to have been generated, at least in part, by partial melting of young and hot subducting oceanic lithospheres<sup>4,6,16,20</sup>. Progressive cooling of the Earth has led to the subduction of older and cooler slabs which, on dehydration, produce fluids that provoke partial melting of the mantle wedge. This generally results in basaltic addition to the continental crust. However, the results from the AVZ suggest that adakites may have been added to the continental crust throughout Earth's history wherever a sufficiently young and hot oceanic crust was subducted. □

Received 8 May 1997; accepted 11 May 1998.

- Gill, J. B. *Organic Andesites and Plate Tectonics* (Springer, New York, 1981).
- McDermott, F. & Hawkesworth, C. Th, Pb and Sr isotope variations in young island arc volcanics and oceanic sediments. *Earth Planet. Sci. Lett.* **104**, 1–15 (1991).
- Condomines, M. & Sigmarsson, O. Why are so many magmas close to  $^{238}\text{U}$ - $^{230}\text{Th}$  radioactive equilibrium? *Geochim. Cosmochim. Acta* **57**, 4491–4497 (1993).
- Defant, M. J. & Drummond, M. S. Derivation of some modern arc magmas by melting of young subducted lithosphere. *Nature* **347**, 662–665 (1990).
- Klein, E. M. & Karsten, J. L. Ocean-ridge basalts with convergent-margin geochemical affinities from the Chile Ridge. *Nature* **374**, 52–57 (1995).
- Martin, H. in *The Archaean Crustal Evolution* (ed. Condie, K. C.) 205–259 (Elsevier, Amsterdam, 1995).
- Allegre, C. J. & Condomines, M. Basalt genesis and mantle structure studied through Th isotope geochemistry. *Nature* **299**, 21–24 (1982).
- Cande, S. C. & Leslie, R. B. Late Cenozoic tectonics of the Southern Chile trench. *J. Geophys. Res.* **91**, 471–496 (1986).
- Forsythe, R. D. et al. Pliocene near-trench magmatism in southern Chile: a possible manifestation of ridge collision. *Geology* **14**, 23–27 (1986).
- Lagabriele, Y., Le Moigne, J., Maury, R. C., Cotten, J. & Bourgois, J. Volcanic record of the subduction of an active spreading ridge: Taitao Peninsula (southern Chile). *Geology* **22**, 515–518 (1994).
- Bourgois, J., Martin, H., Lagabriele, Y., Le Moigne, J. & Frutos-Jara, J. Subduction erosion related to spreading-ridge subduction: Taitao Peninsula (Chile margin triple junction area). *Geology* **24**, 723–726 (1996).
- Forsyth, D. W. Fault plane solution and tectonics of the South Atlantic and Scotia Sea. *J. Geophys. Res.* **80**, 1429–1443 (1975).
- Stern, C. R. & Kilian, R. Role of the subducted slab, mantle wedge and continental crust in the generation of adakites from the Andean Austral Volcanic Zone. *Contrib. Mineral. Petrol.* **123**, 263–281 (1996).
- Hildreth, W. & Moorbath, S. Crustal contribution to arc magmatism in the Andes of Central Chile. *Contrib. Mineral. Petrol.* **98**, 455–489 (1988).
- Stern, C. R., Futa, K. & Muehlenbachs, K. in *Andean Magmatism: Chemical and Isotopic Constraints* (eds Harmon, R. S. & Barreiro, B.) 31–46 (Shiva, Cheshire, 1984).
- Martin, H. Archaean and modern granitoids as indicators of changes in geodynamic processes. *Rev. Bras. Geocienc.* **17**, 360–365 (1987).
- Sigmarsson, O., Condomines, M. & Fourcade, S. A detailed Th, Sr and O isotope study of Hekla: differentiation processes in an Icelandic volcano. *Contrib. Mineral. Petrol.* **112**, 20–34 (1992).
- Sigmarsson, O., Condomines, M., Morris, J. D. & Harmon, R. S. Uranium and  $^{10}\text{Be}$  enrichments by fluids in Andean arc magmas. *Nature* **346**, 163–165 (1990).
- Peacock, S. M., Rushmer, T. & Thompson, A. B. Partial melting of subducted oceanic crust. *Earth Planet. Sci. Lett.* **121**, 227–244 (1994).
- Rapp, R. P. & Watson, E. B. Dehydration melting of metabasalt at 8–32 kbar: implications for continental growth and crust-mantle recycling. *J. Petrol.* **36**, 891–931 (1995).
- Schiano, P. et al. Hydrous, silica-rich melts in the sub-arc mantle and their relationship with erupted arc lavas. *Nature* **377**, 595–600 (1995).
- Beattie, P. Uranium-thorium disequilibria and partitioning on melting garnet peridotite. *Nature* **363**, 63–65 (1993).
- LaTourette, T. K., Kennedy, A. K. & Wasserburg, G. J. Thorium-uranium fractionation by garnet: evidence for a deep source and rapid rise of oceanic basalts. *Science* **261**, 739–742 (1993).
- Johnson, K. T. M. Experimental cpx/ and garnet/melt partitioning of REE and other trace elements at high pressures: petrogenetic implications. *Mineral. Mag.* **58**, 454–455 (1994).
- Tribuzio, R., Messiga, B., Vannucci, R. & Bottazzi, P. Rare earth element redistribution during high-pressure-low-temperature metamorphism in ophiolitic Fe-gabbros (Liguria, northwestern Italy): implications for light REE mobility in subduction zone. *Geology* **24**, 711–714 (1996).
- Mezger, K., Hanson, G. N. & Bohlen, S. R. High-precision U-Pb ages of metamorphic rutile: application to the cooling history of high-grade terranes. *Earth Planet. Sci. Lett.* **96**, 106–118 (1989).
- Brenan, J. M., Shaw, H. F., Phinney, D. L. & Ryerson, F. J. Rutile-aqueous fluid partitioning of Nb, Ta, Hf, Zr, U and Th: implications for high field strength element depletions in island-arc basalts. *Earth Planet. Sci. Lett.* **128**, 327–339 (1994).
- Kelmen, P. M. Genesis of high Mg# andesites and the continental crust. *Contrib. Mineral. Petrol.* **120**, 1–19 (1995).
- Leeman, W. P., Smith, D. R., Hildreth, W., Palacz, Z. & Rogers, N. Compositional diversity of late Cenozoic basalts in a transect across the southern Washington Cascades: implications for subduction zone magmatism. *J. Geophys. Res.* **95**, 19561–19582 (1990).

**Acknowledgements.** We thank C. Stern and R. Kilian for the samples, M. Condomines, J. Gill, A. Hochstaedter, M. Schmidt and D. Vielzeuf for discussions, and J. Morris and S. Turner for reviews.

Correspondence and requests for materials should be addressed to O.S. (e-mail: olgeir@opgc.univ-bpclermont.fr).

## A complete primitive rhizodont from Australia

Zerina Johanson\* & Per E. Ahlberg†

\* Palaeontology Section, Australian Museum, 6 College Street, Sydney South, New South Wales 2000, Australia

† Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Studies of the origin<sup>1–3</sup> and developmental genetics<sup>4–7</sup> of tetrapod limbs have focused attention on the need to identify the precise type of sarcopterygian (lobe-finned fish) fin from which limbs evolved. This can only be achieved through a phylogenetic analysis of sarcopterygians. Sarcopterygian fin skeletons vary in structure<sup>8,9</sup>; use of an inappropriate fin skeleton as a model limb precursor will lead to erroneous inferences about the evolution of morphology and the developmental pathways at the fish–tetrapod transition. The pectoral fin of the rhizodont sarcopterygian *Sauripteris* is strikingly limb-like and features prominently in discussions about the origin of limbs<sup>3,7,10–14</sup>. It is thus important to establish the phylogenetic position of rhizodonts. However, their anatomy is incompletely known<sup>15,16</sup>. Published phylogenetic analyses are based on poorly substantiated characters, such as the alleged presence of two external nostrils in the Australian genus *Barameda*<sup>17,18</sup>. Here we present, from the Upper Devonian period of Canowindra, Australia, the most primitive and by far the most complete rhizodont discovered so far. It has a single external nostril but possesses no other derived tetrapod-like features. Our new evidence shows that rhizodonts are more remote from tetrapods than are osteolepiform<sup>18</sup> and elpistostegid<sup>19</sup> lobe-fish. Similarities between rhizodont fins and tetrapod limbs are thus probably convergent, and the pectoral fin of *Sauripteris* should not be used as a model limb precursor.

Sarcopterygii Romer 1955

Rhizodontida Andrews and Westoll 1970

*Gooloogongia* gen. nov.

**Type species.** *Gooloogongia loomesi* sp. nov.

**Diagnosis.** Recognizable as a rhizodont by the possession of posteriorly pointed postparietals<sup>15,17</sup>, short dorsal blade of the cleithrum<sup>15</sup>, long basal lepidotrichial segments in the pectoral fin<sup>15</sup>, doubled dorsal lateral line canal<sup>15</sup>, supratemporal–extratemporal contact<sup>17</sup>, and a very short anterior margin of the median extrascapular<sup>17</sup>. However, it lacks several derived features of other rhizodonts (premaxillary fang, posterior depressed lamina of cleithrum, long basal lepidotrichial segments in posterior fins<sup>15–17</sup>), indicating that it represents the sister group of previously known Rhizodontida. The body length is roughly 80 cm. The very slender recumbent mandibular fangs seem to be autapomorphic.

*Gooloogongia loomesi* sp. nov.

**Etymology.** *Gooloogongia*: after the town of Gooloogong; *loomesi*: in honour of Bruce Loomes, an enthusiastic supporter of the Canowindra 'Age of Fishes' project.

**Holotype.** AMF 96860A and B, a skull and partial lower jaw, incomplete cleithrum, incomplete pectoral fin and body scales, exposed in dorsal view. All the material consists of natural moulds in sandstone; specimen numbers refer to permanent resin casts taken from these moulds, housed at the Australian Museum, Sydney.

**Referred material.** AMF 99900A–D, 99901, 99902, 100073A and B, 100074, 100075 and 101453.

**Locality.** On the road between Canowindra and Gooloogong, New South Wales, Australia. The site is known as Canowindra.

**Horizon.** Mandagery Sandstone, Upper Devonian, Famennian<sup>20</sup>.

**Diagnosis.** As for genus.

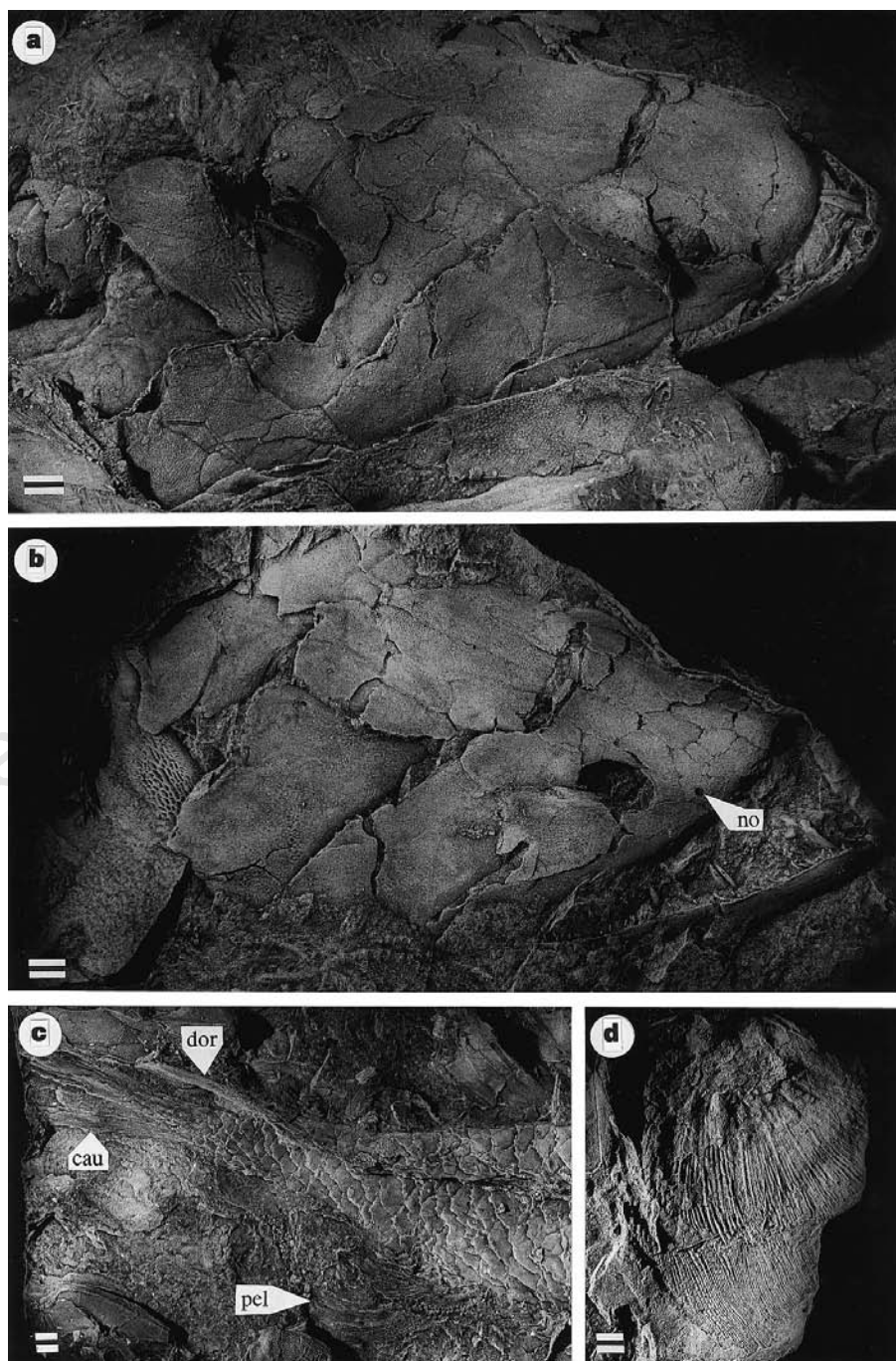
The skull roof of *Gooloogongia* (Figs 1a, b and 2f) is similar to that of *Barameda* (Fig. 2d), the only other rhizodont where it is preserved in articulation<sup>17</sup>. However, it appears less derived than *Barameda* in having several bones in the nasal series. The shared characters of the two genera are probably primitive for the Rhizodontida.

Two nostrils have been reconstructed in *Barameda* (Fig. 2e)<sup>17</sup>. *Gooloogongia* has only one nostril (Figs 1a, b and 2f, g), with the same relationship to surrounding bones as in osteolepiforms<sup>19</sup> and elpistostegids (panderichthyids)<sup>21,22</sup>, which are members of the tetrapod stem group (Fig. 3). Re-examination of the *Barameda* material indicates that it too conforms to this pattern. In osteolepi-

forms, elpistostegids and tetrapods, the single external nostril is associated with a palatal nostril or choana. Unfortunately the palate is not preserved in *Gooloogongia*.

The cheek, previously largely unknown in rhizodonts<sup>15,17</sup> (Fig. 2e), is osteolepiform-like in *Gooloogongia* (Figs 1a, b and 2g). However, it has a broad preopercular (Pop), which is an apparently primitive feature that is shared with porolepiforms and onychodonts<sup>16,18</sup>, but is absent in osteolepiforms and tetrapods<sup>16,18</sup> where this bone is narrow.

The preserved elements of the pectoral fin endoskeleton (Figs 1d, 4b) resemble those of *Sauripteris*<sup>8,14</sup> and other rhizodonts<sup>8,17</sup> (Fig. 4a), indicating that it was probably of the same elaborate

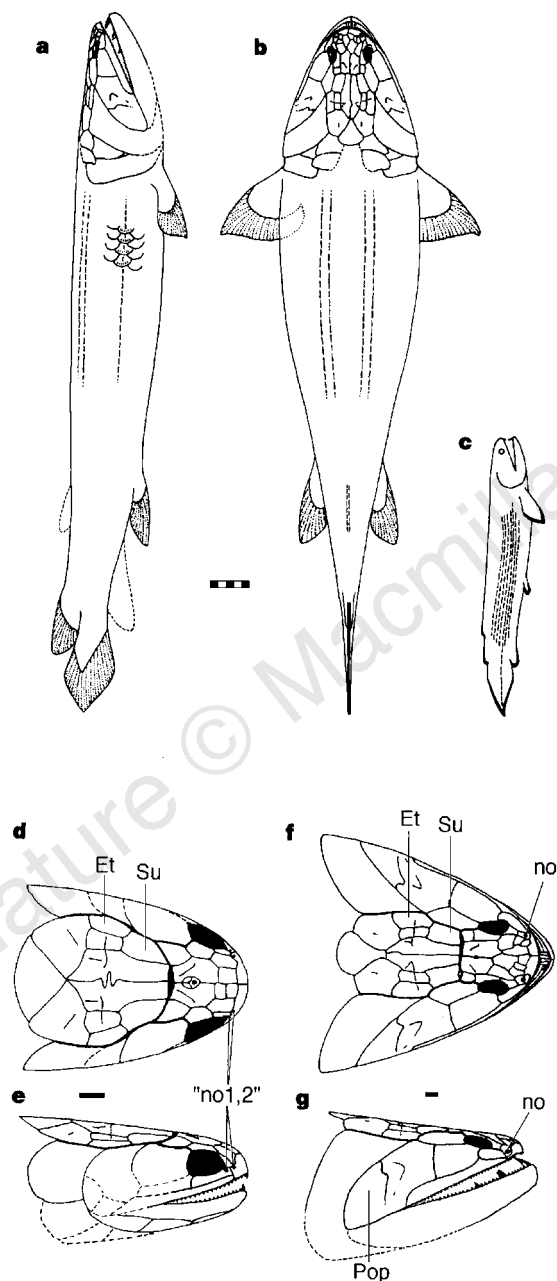


**Figure 1** Specimens AMF 9680, 100073 and 99900 of *Gooloogongia loomesi*. **a**, AMF 96860, holotype of *Gooloogongia loomesi*, a skull in dorsolateral view. The large notch in the posterior margin of the opercular is an artefact. **b**, AMF 100073, flattened skull in dorsolateral view, showing the single external nostril (no). Both of these skulls have protruding lower jaws with slender recumbent fangs. **c**, AMF

99900, posterior part of body showing caudal (cau), second dorsal (dor) and pelvic (pel) fins. The basal portions of the lepidotrichia are short, as in most other osteichthyans. **d**, AMF 99900, pectoral fin in ventral view, showing long basal portions of lepidotrichia and traces of a complex endoskeleton. Scale bars, 1 cm.

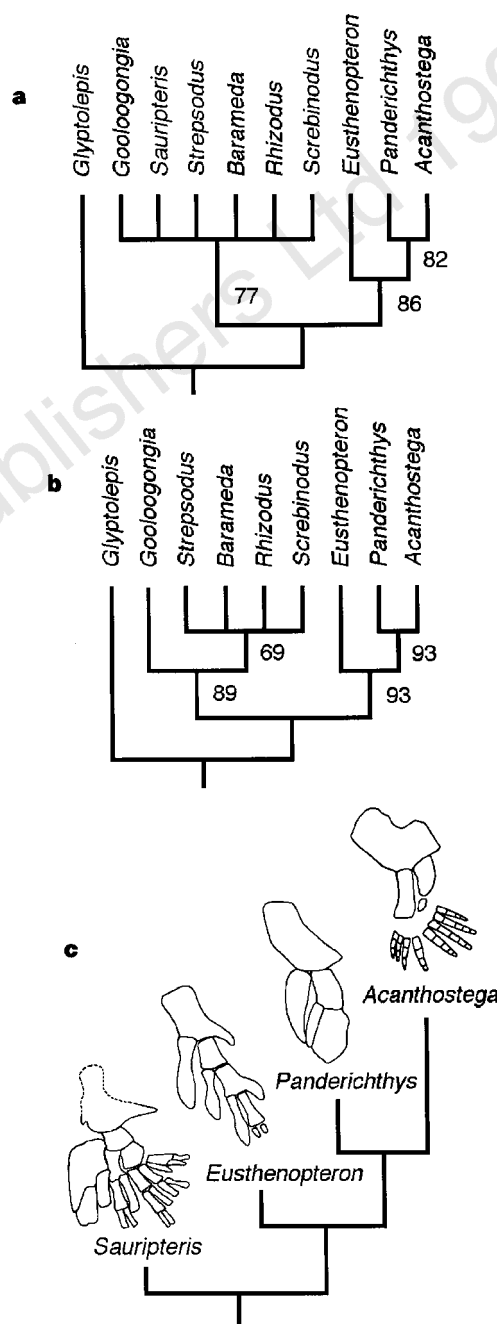
pattern. The pectoral fin skeletons of osteolepiforms and elpistostegids are much simpler (Fig. 3).

The pectoral fin lepidotrichia of *Gooloogongia* have basal segments that are equal in length to the jointed distal portions (Figs 1d and 4b). In other rhizodonts the basal segments are much longer than the jointed parts<sup>15</sup>, whereas the generalized osteichthyan condition (see in, for example, actinopterygians and osteolepiforms) is to have very short basal segments<sup>15</sup>. The posterior fins of *Gooloogongia* show this primitive condition (Fig. 1c), whereas other rhizodonts have long basal lepidotrichial segments in all fins<sup>15</sup>.



**Figure 2** Reconstructions of *Gooloogongia loomesi*, *Strepsodus ancunamensis* and *Barameda*. **a, b**, Lateral and dorsal reconstructions of *Gooloogongia loomesi*. A sample area of squamation is shown on the flank, with preserved parts of the lateral line canals (dashed lines). Scale bar, 5 cm. **c**, *Strepsodus ancunamensis*, lateral view (modified from ref. 15). A probable juvenile, drawn to scale with **a, b**. Note multiple lateral line canals. **d, e**, Skull of *Barameda decipiens* in dorsal and lateral views (from ref. 17), showing two reconstructed nostrils ("no 1, 2"). Dashed lines are hypothetical. Scale bar, 1 cm. **f, g**, Skull of *Gooloogongia loomesi* in dorsal and lateral views. Pop, preopercular; Et, extratemporal; Su, supratemporal.

The body form of *Gooloogongia*, with its broad head, dorsally facing eyes and small caudal fin (Fig. 2a, b) indicates that it may have been a benthic sit-and-wait predator. It differs considerably from the only other known rhizodont body, of the Carboniferous genus *Strepsodus*<sup>15</sup> (Fig. 2c), which has a symmetrical tail and reduced



**Figure 3** Relationships between rhizodonts, *Eusthenopteron*, *Panderichthys* tetrapods. **a, b**, Cladograms generated by PAUP3.1 (exhaustive search) from 29 morphological characters, showing relationships of rhizodonts, *Eusthenopteron*, *Panderichthys* and tetrapods. The outgroup is the porolepiform *Glyptolepis*. Figures at internodes are bootstrap support values (heuristic, 100 replicates). **a**, Includes *Sauripteris*, which can only be scored for pectoral fin characters and thus lowers the phylogenetic resolution. Strict consensus of 135 trees, with length 40; consistency index, 0.825; homoplasy index, 0.175; retention index, 0.848, and rescaled consistency index, 0.699. **b**, Excludes *Sauripteris*, and resolves *Gooloogongia* as the sister group of other Rhizodontida. Strict consensus of 15 trees; length, 39; consistency index, 0.846; homoplasy index, 0.154; retention index, 0.864; and rescaled consistency index, 0.731. **c**, Pectoral appendages mapped onto phylogeny.



median fins. Superficially *Strepsodus* resembles an elpistostegid or basal tetrapod, whereas *Gooloogongia* is closer to the generalized sarcopterygian condition exemplified by *Osteolepis*<sup>19</sup>. *Strepsodus* has multiple lateral lines<sup>15</sup>; *Gooloogongia* shows a simpler pattern with a double dorsolateral line as its only elaboration.

A cladistic analysis (Fig. 3) places the Rhizodontida as the most basal member of the tetrapod stem group, with high bootstrap support. *Gooloogongia* seems to be the most primitive rhizodont (Fig. 3b), although this resolution is lost when *Sauripteris* (which

can only be scored for pectoral fin characters) is included (Fig. 3a).

The main puzzle of rhizodont relationships to date has been the apparent contradiction between an elaborate limb-like pectoral fin skeleton<sup>3,7,8,10–14</sup> and the supposed presence of two external nostrils<sup>15,17</sup>. Elpistostegids and osteolepiforms (represented here by *Panderichthys* and *Eusthenopteron*) have the opposite character complement: a tetrapod-like nasal region with a single external nostril and a choana, but a simple pectoral fin skeleton (Fig. 3) that contains no elements resembling digits<sup>16–18,21,23</sup>. *Gooloogongia* shows that rhizodonts have a single external nostril, thus resolving this spurious character conflict. However, it exhibits no other tetrapod-like characters.

The rhizodont skull morphology illustrated by *Gooloogongia* and *Barameda* differ from the elpistostegid and tetrapod pattern in lacking frontal bones and having a short snout, widely separated eyes, a well developed intracranial joint between parietals and postparietals, and extratemporals (Fig. 4c–e). These characteristics of rhizodont skulls are shared with porolepiforms<sup>19</sup>, which are basal stem-group lungfishes<sup>16,18</sup>; these characteristics are therefore probably shared primitive characters of the lungfish and tetrapod stem groups. They also occur in osteolepidid (but not tristichopterid) osteolepiforms<sup>19</sup>, which may indicate that osteolepiforms are paraphyletic relative to elpistostegids and tetrapods.

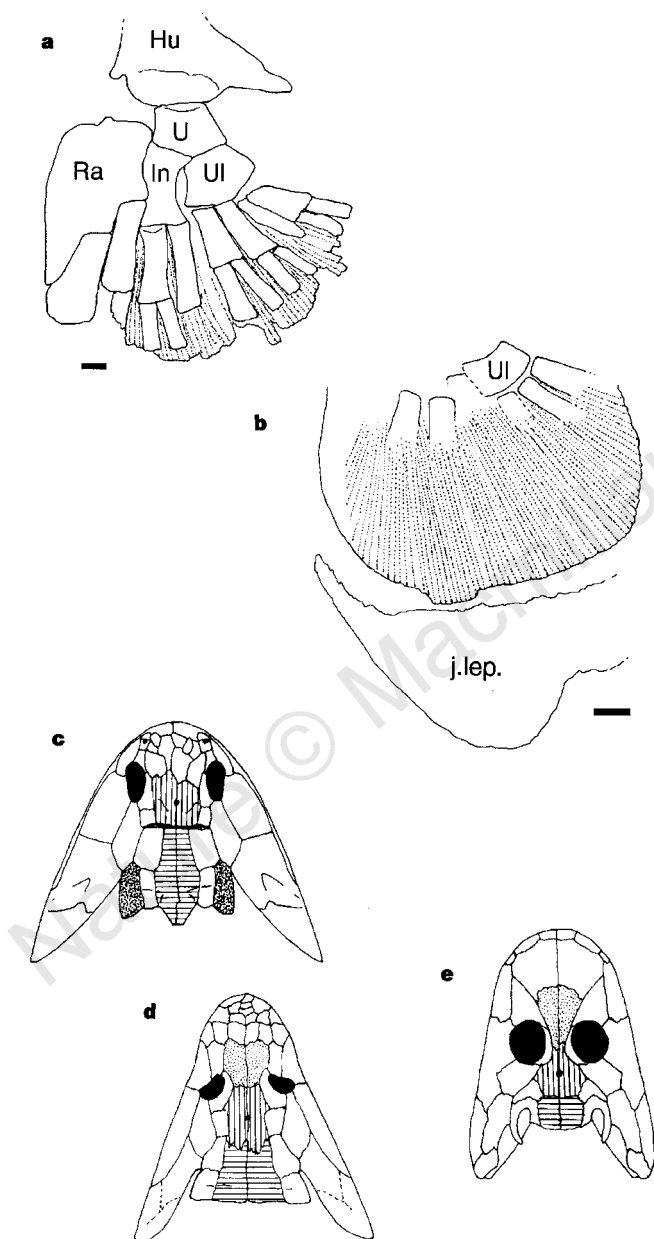
The description of *Gooloogongia* improves our understanding of rhizodont anatomy, and shows conclusively that rhizodonts are less closely related to tetrapods than are osteolepiforms and elpistostegids. We conclude that the similarities between the pectoral appendage skeletons of rhizodonts and tetrapods<sup>7,14</sup> are convergent (Fig. 3), and urge that rhizodont pectoral fins not be used as model ancestors for tetrapod limbs. □

## Methods

**Phylogenetic analysis.** The analysis was performed using the software package PAUP with a data matrix (see Supplementary Information) of 10 taxa (the porolepiform *Glyptolepis*<sup>9,18,19</sup>, the rhizodonts *Gooloogongia*, *Barameda*<sup>17</sup>, *Sauripteris*<sup>8,14</sup>, *Strepsodus*<sup>8,15</sup>, *Rhizodus*<sup>8,15</sup> and *Screbiodus*<sup>8,15</sup>, the tristichopterid osteolepiform *Eusthenopteron*<sup>19</sup>, the elpistostegid *Panderichthys*<sup>21</sup> and the tetrapod *Acanthostega*<sup>2,24</sup>) scored for 29 morphological characters (see Supplementary information). Most parsimonious trees were identified using the exhaustive search algorithm. All characters were unordered and weighted equally. Bootstrap values were determined from 100 replicates using the heuristic algorithm.

Received 3 December 1997; accepted 24 April 1998.

- Coates, M. I. & Clack, J. A. Polydactyly in the earliest known tetrapod limbs. *Nature* **347**, 66–69 (1990).
- Coates, M. I. The Devonian tetrapod *Acanthostega gunnari* Jarvik: postcranial anatomy, basal tetrapod interrelationships and patterns of skeletal evolution. *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
- Shubin, N. The evolution of paired fins and the origin of tetrapod limbs. *Evol. Biol.* **28**, 39–85 (1995).
- Shubin, N. H. & Alberch, P. A morphogenetic approach to the origin and basic organization of the tetrapod limb. *Evol. Biol.* **20**, 319–387.
- Coates, M. I. in *Developmental Patterning of the Vertebrate Limb* (eds Hinchliffe, J. R., Hurle, J. M. & Summerbell, D.) 325–337 (Plenum, New York, 1991).
- Sordino, P., van der Hoeven, F. & Duboule, D. *Hox* gene expression in teleost fins and the origin of vertebrate digits. *Nature* **375**, 678–681 (1995).
- Shubin, N., Tabin, C. & Carroll, S. Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
- Andrews, S. M. & Westoll, T. S. The postcranial skeleton of rhipidistian fishes excluding *Eusthenopteron*. *Trans. R. Soc. Edinb.* **68**, 391–489 (1970).
- Ahlberg, P. E. Paired fin skeletons and relationships of the fossil group Porolepiformes (Osteichthyes: Sarcopterygii). *Zool. J. Linn. Soc.* **96**, 119–166 (1989).
- Gregory, W. K. The limbs of *Eryops* and the origin of paired limbs from fins. *Science* **33**, 508–509 (1911).
- Broom, R. On the origin of the cheiropterygium. *Bull. Am. Mus. Nat. Hist.* **32**, 459–464 (1913).
- Romer, A. S. *The Vertebrate Body* (W. B. Saunders, Philadelphia, 1949).
- Schmalhausen, I. I. *The Origin of Terrestrial Vertebrates*. (Academic, New York, 1968).
- Daeschler, E. & Shubin, N. Fish with fingers? *Nature* **391**, 133 (1998).
- Andrews, S. M. Rhizodont crossospterygian fish from the Dinantian of Foulden, Berwickshire, Scotland, with a re-evaluation of this group. *Trans. R. Soc. Edinb. Earth Sci.* **76**, 67–95 (1985).
- Cloutier, R. & Ahlberg, P. E. in *Interrelationships of Fishes* (eds Stiassny, M. L. J., Parenti, L. R. & Johnson, G. D.) 445–479 (Academic, San Diego, 1996).
- Long, J. A. A new rhizodontiform fish from the Early Carboniferous of Victoria, Australia, with remarks on the phylogenetic position of the group. *J. Vert. Paleontol.* **9**, 1–17 (1989).
- Ahlberg, P. E. A re-examination of sarcopterygian interrelationships, with special reference to the Porolepiformes. *Zool. J. Linn. Soc.* **103**, 241–287 (1991).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. 1 (Academic, New York, 1980).



**Figure 4** Morphology of ?*Sauripteris*, *Gooloogongia*, *Panderichthys* and *Acanthostega*. **a**, Pectoral fin skeleton of ?*Sauripteris* (from ref. 14), showing humerus (Hu), radius (Ra), ulna (U), intermedium (In), ulnare (Ul), branching distal radials and long unjointed basal segments of lepidotrichia (dotted lines). Left is anterior. Scale bar, 1 cm. **b**, Pectoral fin skeleton of *Gooloogongia*, AMF 99900, same orientation, showing large jointed lepidotrichial web (j.lep.). Scale bar, 1 cm. **c–e**, Heads of the rhizodont *Gooloogongia* (**c**), elpistostegid *Panderichthys* (**d**) (modified from ref. 21) and tetrapod *Acanthostega* (from ref. 24) (**e**), showing frontals (light shading), extratemporals (dark shading), parietals (vertical hatching) and postparietals (horizontal hatching). In *Gooloogongia* and *Panderichthys*, operculars and extrascapulars have been omitted.

20. Young, G. C. in *Palaeozoic Vertebrate Biostratigraphy and Biogeography* (ed. Long, J. A.) 208–251 (Belhaven, London, 1993).
21. Vorobyeva, E. I. & Schultze, H.-P. in *Origins of the Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ., Ithaca, NY, 1991).
22. Schultze, H.-P. & Arsenault, M. The panderichthys fish *Elpistostege*: a close relative of tetrapods? *Paleontology* **28**, 293–309 (1985).
23. Schultze, H.-P. Dipnoans as sarcopterygians. *J. Morphol.* **1** (suppl.), 39–74 (1986).
24. Clack, J. A. *Acanthostega gunnari*, a Devonian tetrapod from Greenland; the snout, palate and ventral parts of the braincase, with a discussion of their significance. *Meddr. Grønland Geoscience* **31**, 1–24 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank the Australian Museum for their award of a Visiting Fellowship to P.E.A. A. Ritchie for inviting us to work on the Canowindra material and J. Long and J. Jeffery for discussions and access to rhizodont material. Z.J. thanks J. Fairfax for financial support. This Letter is dedicated to the memory of S. M. Andrews.

## Selfish genes: a green beard in the red fire ant

Laurent Keller\* & Kenneth G. Ross†

\* Institute of Zoology and Animal Ecology, University of Lausanne, Bâtiment de Biologie, 1015 Lausanne, Switzerland

† Department of Entomology, University of Georgia, Athens, Georgia 30602-2603, USA

A 'green-beard' gene is defined as a gene that causes a phenotypic effect (such as the presence of a green beard or any other conspicuous feature), allows the bearer of this feature to recognize it in other individuals, and causes the bearer to behave differently towards other individuals depending on whether or not they possess the feature<sup>1–3</sup>. Such genes have been proposed on theoretical grounds to be agents mediating both altruism and intragenomic conflicts<sup>1,2</sup>, but until now few, if any, of these genes have been identified<sup>4,5</sup>. Here we provide evidence of a green-beard gene in the red imported fire ant, *Solenopsis invicta*. In polygyne (multiple-queen) colonies, all egg-laying queens are *Bb* heterozygotes at the locus *Gp-9* (ref. 6). Previous studies suggested that *bb* females die prematurely from intrinsic causes<sup>6</sup>; we now show that *BB* queens initiating reproduction are killed by workers, and that it is primarily *Bb* rather than *BB* workers that are responsible for these executions. This implies that allele *Gp-9<sup>b</sup>* is linked to a green-beard allele that preferentially induces workers bearing the allele to kill all queens that do not bear it. Workers appear to distinguish *BB* from *Bb* queens on the basis of a transferable odour cue.

We mimicked the natural recruitment of new reproductive queens into polygyne nests by reintroducing young queens into their parental colonies after they had been kept in small colony fragments with workers but not other queens for three days. (The absence of mature reproductive queens induces reproductive development of young queens<sup>7</sup>.) There was a strong association between queen genotype at *Gp-9* and the probability of being attacked: all attacked queens were homozygous for the *B* allele, whereas none of the queens with a copy of allele *b* faced significant worker aggression (Table 1). A separate experiment in which worker attacks were not interrupted showed that such attacks invariably led to the death of a queen within 15 minutes ( $n = 50$ ).

We next compared the genotypes of workers attacking *BB* queens with those of workers sampled randomly from the same population and discovered that attackers were much more likely to carry the *b* allele (82.5 versus 59.3% of genotypes;  $P < 0.01$ ). To confirm this, we did a second experiment in which we compared the genotypes of workers attacking *BB* queens with those of workers in the vicinity of non-attacked (*Bb*) queens in the same colonies. The proportion of *Bb* and *bb* workers surrounding attacked (*BB*) queens was significantly higher than the proportion surrounding non-attacked (*Bb*)

queens (Table 2). Our assay may considerably underestimate the true extent of genotypic bias among workers attacking *BB* queens because these attacks elicited the formation of compact worker groups around the queens, making it impossible to collect only attacking workers. Thus, although our experiments demonstrate that such attacks are undertaken primarily by workers with at least one copy of allele *Gp-9<sup>b</sup>*, they do not allow us to determine whether the attacks are carried out only by these workers. The possibility that these results are due to workers with the *b* allele generally having a lower threshold for aggression can be excluded, because such individuals were not overrepresented among the workers attacking foreign heterospecific ant workers (*Aphaenogaster* sp.) introduced into nests ( $n = 788$ ;  $G = 2.42$ ; d.f. = 1;  $P = 0.12$ ; workers with the *b* allele actually were underrepresented among these attackers).

Some of the workers involved in attacks on *BB* queens subsequently were attacked by nestmates, suggesting that they might have acquired a distinctive odour from the attacked queens. To test this hypothesis, we rubbed randomly chosen workers against the cuticle of *BB* or *Bb* queens and then placed them in groups of nestmate

**Table 1** Proportion of young queens of each *Gp-9* genotype attacked by workers

| Age                                | Proportion of queens attacked |                       |                      |
|------------------------------------|-------------------------------|-----------------------|----------------------|
|                                    | <i>Gp-9</i> genotype          |                       |                      |
|                                    | <i>BB</i>                     | <i>Bb</i>             | <i>bb</i>            |
| 7–10 days<br>( $n = 19$ colonies)  | 0.61<br>( $n = 90$ )          | 0.00<br>( $n = 275$ ) | 0.00<br>( $n = 5$ )  |
| 11–14 days<br>( $n = 18$ colonies) | 0.91<br>( $n = 11$ )          | 0.00<br>( $n = 327$ ) | 0.00<br>( $n = 11$ ) |

The frequencies of attacks on 7–10-day-old queens varied significantly according to genotype ( $G = 190.78$ ; d.f. = 2;  $P < 0.0001$ ). The proportion of *BB* queens attacked was significantly greater than the proportion of either *Bb* queens ( $G = 189.16$ ; d.f. = 1;  $P < 0.0001$ ) or *bb* queens ( $G = 9.04$ ; d.f. = 1;  $P = 0.003$ ) attacked (the few *bb* queens found presumably had not yet succumbed to the age-dependent lethal effects of this genotype<sup>6</sup>). The same pattern was revealed within individual colonies: in each of the 16 colonies in which *BB* queens were present, the proportion of such queens attacked was significantly greater than the proportion of attacked queens with the other two genotypes (binomial probability,  $P < 0.001$ ). A similar association between queen genotype and aggression occurred for 11–14-day-old queens ( $G = 84.06$ ; d.f. = 2;  $P < 0.0001$ ), with the proportion of *BB* queens attacked again being significantly greater than the proportion of either *Bb* queens ( $G = 83.41$ ; d.f. = 1;  $P < 0.0001$ ) or *bb* queens ( $G = 23.61$ ; d.f. = 1;  $P < 0.001$ ) attacked. At the colony level, the proportion of attacked *BB* queens of this older class again was greater than the proportion of attacked queens with the other two genotypes in each of the three colonies in which *BB* queens were present. Two lines of evidence suggest that *BB* queens are killed as they approach sexual maturity and become potential egg layers (about 10 d after adult emergence<sup>8,20</sup>). First, the proportion of attacked *BB* queens was higher for 11–14-day-old queens than for 7–10 day old queens ( $G = 4.58$ ; d.f. = 1;  $P = 0.03$ ). Second, among queens in the younger age class, the non-attacked queens with genotype *BB* were significantly lighter ( $12.8 \pm 1.3$  mg;  $n = 35$ ) than the attacked queens with this genotype ( $14.1 \pm 1.1$  mg;  $n = 55$ ; two-way ANOVA, weight difference:  $F = 23.55$ ;  $P < 0.0001$ ; colony effect:  $F = 1.95$ ;  $P < 0.05$ ; interaction:  $F = 0.831$ , NS). Because weight is a good indicator of the age of maturing queens<sup>20</sup>, these data suggest that younger and lighter *BB* queens were attacked relatively less frequently. Such age-associated attacks on *BB* queens may explain the decrease in the proportion of *BB* genotypes among queens as they age: this proportion was 0.24 in the 7–10-day-old queens, and 0.03 in the 11–14-day-old queens ( $G = 77.69$ ; d.f. = 1;  $P < 0.001$ ).

**Table 2** Number of workers of each *Gp-9* genotype surrounding attacked, *Gp-9<sup>BB</sup>* queens and non-attacked, *Gp-9<sup>Bb</sup>* queens

| Worker<br><i>Gp-9</i> genotype | Queen<br><i>Gp-9</i> genotype |                             |
|--------------------------------|-------------------------------|-----------------------------|
|                                | <i>BB</i><br>(attacked)       | <i>Bb</i><br>(non-attacked) |
| <i>BB</i>                      | 50 (0.213)                    | 81 (0.344)                  |
| <i>Bb</i>                      | 184 (0.783)                   | 147 (0.626)                 |
| <i>bb</i>                      | 1 (0.004)                     | 7 (0.030)                   |
| Total                          | 235                           | 235                         |

Proportions of attacking and non-attacking workers with each genotype are shown in parentheses. There was a significant association between queen and worker genotypes ( $G = 16.61$ ; d.f. = 2;  $P = 0.0002$ ), with attacks on *BB* queens being made preferentially by workers having the *b* allele. The difference remains highly significant, both when *bb* workers are eliminated from the analysis ( $G = 11.47$ ; d.f. = 1;  $P < 0.001$ ) and when they are pooled with *Bb* workers ( $G = 10.24$ ; d.f. = 1;  $P = 0.001$ ). The same pattern was found within individual colonies. Of the nine colonies that contained three or more queens of each genotype, eight had a relative overrepresentation of workers with allele *b* attacking queens, an outcome that departs significantly from the null expectation that 50% of nests should have such overrepresentations (binomial test,  $P < 0.02$ ). Furthermore, the overrepresentation of allele *b* in attacking workers was significant in two of these colonies ( $G = 6.89$ ; d.f. = 1;  $P < 0.01$  and  $G = 5.07$ ; d.f. = 1;  $P = 0.02$ ). The few *bb* workers found presumably were very young workers that had not yet succumbed to the age-dependent lethal effects of this genotype<sup>6</sup>.

workers. Those rubbed against *BB* queens elicited significantly higher levels of aggression (aggression level was  $2.5 \pm 0.7$  ( $n = 10$ ); Mann–Whitney *U* test,  $Z = 3.71$ ,  $P = 0.0002$ ) and were killed significantly more often (40% ( $n = 10$ );  $G = 6.56$ , d.f. = 1,  $P = 0.01$ ) than those rubbed against *Bb* queens (aggression level,  $0.8 \pm 0.4$  ( $n = 10$ ), 0% killed ( $n = 10$ )). These data, together with the finding that *BB* queens are attacked when they attain sexual maturity (Table 1), suggest that recognition and selective elimination of *BB* queens may be triggered by two chemical cues, one signalling a queen's sexual maturity and the other her *Gp-9*-linked genotype. Tight coupling between a queen's reproductive state and pheromone production has been demonstrated in *S. invicta*<sup>8</sup>, as has the ability of workers to assess the level of pheromone production by individual queens<sup>9</sup>. Thus, the green-beard allele linked to *Gp-9*<sup>b</sup> may induce workers bearing it to kill all sexually mature queens except those possessing a specific chemical signature encoded by this allele.

The green-beard gene responsible for differences in queen odours and in the aggressive behaviour of workers with different *Gp-9* genotypes may be *Gp-9* itself, or one or more genes in very strong gametic disequilibrium with *Gp-9*. The enzyme-encoding locus *Pgm-3* (ref. 6) is tightly linked to and in strong disequilibrium with *Gp-9*. All females of genotype *Pgm-3*<sup>AA</sup> also have genotype *Gp-9*<sup>BB</sup> (ref. 6), accounting for the strong worker discrimination against *Pgm-3*<sup>AA</sup> queens reported previously for polygyne fire ants<sup>7,10,11</sup>. Analyses considering both genes simultaneously show that all described phenotypic and behavioural variation among queens and workers can be accounted for by *Gp-9* genotype alone in the Georgia population under study (our unpublished results). Moreover, although workers eliminated all 33 *Pgm-3*<sup>AA</sup> queens in the re-introduction experiments (Table 1), they also eliminated 26 *Pgm-3*<sup>Aa</sup> and five *Pgm-3*<sup>aa</sup> queens, indicating that *Pgm-3*<sup>a</sup> is not in complete disequilibrium with the green-beard allele. In contrast, workers eliminated no queens with allele *Gp-9*<sup>b</sup> in these same experiments. Finally, separate experiments show that workers destroy 100% of introduced *Gp-9*<sup>BB</sup> queens when these are fully sexually mature (our unpublished results), which explains the complete absence of egg-laying queens with this genotype in polygyne colonies in Georgia<sup>6</sup>. These data indicate that either *Gp-9* is directly responsible for the effects reported, or that the actual gene(s) involved is in complete disequilibrium with *Gp-9* but not *Pgm-3*.

The mechanism of selection against *Gp-9*<sup>BB</sup> queens is similar to the process of meiotic drive. Driving elements invade a population by distorting the outcomes of meiosis, such that these elements are carried by more than 50% of viable gametes produced by heterozygotes. Allele *Gp-9*<sup>b</sup> causes workers to destroy queens without the allele, biasing allele frequencies in sexuals produced by the colony. Fire-ant colony productivity is limited mostly by worker number rather than queen number<sup>12</sup>, so selective elimination of *BB* queens results in an overall increase in the reproductive success of *Bb* queens and thus in the number of copies of the *b* allele transmitted to subsequent generations. Models show that, all else being equal, an outlaw gene<sup>13</sup>, such as *Gp-9*<sup>b</sup>, that causes the destruction of individuals not bearing it in favour of those that do, should spread rapidly and become fixed in a population<sup>14,15</sup>, yet *Gp-9* is polymorphic in all polygyne *S. invicta* populations studied in South America and the USA (ref. 6, and C. J. DeHeer, D. D. Shoemaker and K.G.R., unpublished results). Fixation of *Gp-9*<sup>b</sup> apparently is prevented primarily because queens (as well as workers) homozygous for this allele die prematurely (that is, the allele behaves as a recessive lethal), although gene flow from a different social form fixed for the alternative allele probably also plays a role<sup>6</sup>.

Our results show that all components of a green-beard effect<sup>3</sup>—a detectable phenotypic feature, the ability to recognize the feature, and different responses towards individuals possessing or not possessing the feature—are present in polygyne *S. invicta* and are mediated by a gene or group of genes closely linked to *Gp-9*. The analogy between the green-beard effect reported here and meiotic

drive lies not only in the manner in which one allele biases its transmission, but also in the mechanism preventing its fixation. In both cases, the advantage of increased transmission is counteracted by negative viability and/or fertility effects of the allele when in the homozygous condition<sup>16,17</sup>. This trade-off may explain why green-beard genes have seldom been reported. In the absence of counter-vailing evolutionary pressures, polymorphisms at green-beard loci are expected to be present only as transient phases in the history of a population and thus will usually go undetected. □

## Methods

**Worker aggression towards queens.** The queens originated from 37 polygyne colonies collected in northern Georgia, USA. Entire colonies were transferred into laboratory rearing units using standard procedures, and the number of reproductive queens in each colony was reduced to four<sup>7</sup>. All sexuals were removed from these colonies except for 40 adult winged queens with unsclerotized cuticles (0–3 days old). All winged queens in 19 of the colonies were removed four days later, and 18–20 of these queens from each colony were placed individually in small fragments of the source colony containing ~300 worker brood and adults but no other queens. After three days of separation, the queens (then 7–10 days old) were returned individually to their parent colony (each containing at least 5,000 workers) and the level of aggression directed towards them was recorded during the 5 min immediately following introduction. An identical procedure was followed with the remaining 18 colonies, except that queens were removed from their parent colony after 8 days (thus they were 11–14 days old when returned). Assessment of the level of aggression was done without knowledge of *Gp-9* genotypes, which were determined later by means of starch-gel electrophoresis<sup>6</sup>. Association between genotype and aggression was determined using *G*-tests<sup>18</sup>. Other experiments (ref. 7, and our unpublished results) showed that worker responses to queens of alternative genotypes are independent of whether queens are introduced sequentially (as in this experiment) or in larger groups.

**Genotypes of workers attacking *Gp-9*<sup>BB</sup> queens and *Aphaenogaster* workers.** The preliminary experiment in which genotype frequencies of attacking workers were compared to population frequencies was done using 10 polygyne colonies collected in northern Georgia. Twenty 10–13-day-old queens from each of these colonies were reintroduced into their parent colonies after having been separated for three days (see above). Attacked queens and the clusters of workers surrounding them were collected with forceps. All 22 queens attacked were found to be *Gp-9*<sup>BB</sup> homozygotes. We determined the *Gp-9* genotypes of 5–11 (average, 8.9) workers surrounding each of these queens, and the average frequencies were estimated using a resampling procedure in which a single worker genotype per colony was drawn at random (with replacement) 1,000 times. These frequencies were compared to genotype frequencies estimated in the source population in the same year (based on 406 workers sampled randomly from 181 colonies; the frequencies and their 99% confidence intervals were obtained using the resampling procedure<sup>6</sup>). The follow-up experiment was conducted with another 19 colonies, the same as were used to determine the level of aggression of workers against 7–10-day-old queens (see above). All 55 queens attacked were found to be *BB* homozygotes (Table 1). We determined the *Gp-9* genotype of five workers surrounding each of 47 of these queens. As a control, we collected and genotyped five workers in the vicinity of each of 47 *Bb* (non-attacked) queens. The same numbers of *BB* and *Bb* queens were used from each colony (maximum four per colony) to control for possible differences in genotype frequencies among colonies. To rule out the possibility that behavioural differences among workers with different *Gp-9* genotypes stem simply from intrinsic differences in their aggressiveness, we sequentially introduced four workers of another ant species (*Aphaenogaster* sp.) into each of 20 colonies and collected five fire ant workers attacking each alien ant. The genotype frequencies of the attacking fire ants were compared to those of 400 randomly sampled workers (20 workers from each of the 20 colonies). This procedure again controlled for possible differences in genotype frequencies among colonies. Associations between genotype frequencies were determined using *G*-tests<sup>18</sup>.

**Queen-odour transfer.** Ten of the 11–14-day-old queens attacked by workers were kept individually for at least 10 min in a small isolation unit before a randomly selected live worker from the same parent colony was rubbed against the queen's thorax and abdomen. The worker was reintroduced into the parent



colony 2 min later and aggression was recorded during the 5 min immediately after introduction. The same procedure was followed for workers that were rubbed against non-attacked queens, with a single such queen being chosen randomly to match by colony each attacked queen. This procedure controlled for possible differences in worker behaviour among colonies. Subsequent genetic analysis revealed that all attacked queens were *Gp-9<sup>BB</sup>* homozygotes, whereas all non-attacked queens were *Gp-9<sup>Bb</sup>* heterozygotes. Levels of aggression were defined as: 0, no aggression; 1, infrequent biting; 2, frequent biting but attacked workers not immobilized; and 3, frequent biting with attacked workers immobilized. Scoring was done without knowledge of whether test workers had been rubbed against attacked (*BB*) or non-attacked (*Bb*) queens. In half of the replicates, we first introduced the worker rubbed against a *Gp-9<sup>BB</sup>* queen and in the other half the worker rubbed against a *Gp-9<sup>Bb</sup>* queen. This procedure controlled for possible changes in workers' behaviour in recipient colonies through time.

Received 7 November 1997; accepted 29 May 1998.

- Hamilton, W. D. The genetical evolution of social behavior. 2. *J. Theor. Biol.* **7**, 17–52 (1964).
- Dawkins, R. *The Selfish Gene* (Oxford University Press, New York, 1976).
- Haig, D. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 4th edn 284–304 (Blackwell, Oxford, 1997).
- Queller, D. J. C. Kin selection and frequency dependence: a game theoretical approach. *Biol. J. Linn. Soc.* **23**, 133–143 (1984).
- Haig, D. Gestational drive and the green-bearded placenta. *Proc. Natl Acad. Sci. USA* **93**, 6547–6551 (1996).
- Ross, K. G. Multilocus evolution in fire ants—effects of selection, gene flow and recombination. *Genetics* **145**, 961–974 (1997).
- Keller, L. & Ross, K. G. Phenotypic basis of reproductive success in a social insect: genetic and social determinants. *Science* **260**, 1107–1110 (1993).
- Willer, D. E. & Fletcher, D. J. C. Differences in inhibitory capability among queens of the ant *Solenopsis invicta*. *Physiol. Entomol.* **11**, 475–482 (1986).
- Fletcher, D. J. C. & Blum, M. S. Regulation of queen number by workers in colonies of social insects. *Science* **219**, 312–314 (1983).
- Ross, K. G. Strong selection on a gene that influences reproductive competition in a social insect. *Nature* **355**, 347–349 (1992).
- Ross, K. G., Vargo, E. L. & Keller, L. Simple genetic basis for important social traits in the fire ant *Solenopsis invicta*. *Evolution* **50**, 2387–2399 (1996).
- Vargo, E. L. & Fletcher, D. J. C. On the relationship between queen number and fecundity in polygynous colonies of the fire ant, *Solenopsis invicta*. *Physiol. Entomol.* **14**, 223–232 (1989).
- Ridley, M. & Grafen, A. Are green beard genes outlaws? *Anim. Behav.* **29**, 954–955 (1981).
- Wade, M. J. & Beeman, R. W. The population dynamics of maternal-effect selfish genes. *Genetics* **138**, 1309–1314 (1994).
- Wilson, D. S. & Dugatkin, L. Nepotism vs TFT or why should you be nice to your rotten brother? *Evol. Ecol.* **5**, 291–299 (1991).
- Durand, D., Ardlie, K., Buttel, L., Levin, S. A. & Silver, L. M. Impact of migration and fitness on the stability of lethal *r*-haplotype polymorphism in *Mus musculus*: a computer study. *Genetics* **145**, 1093–1108 (1997).
- Silver, L. M. The peculiar journey of a selfish chromosome: mouse *t* haplotypes and meiotic drive. *Genetics* **9**, 250–254 (1993).
- Sokal, R. R. & Rohlf, F. J. *Biometry. The Principles and Practice of Statistics in Biological Research* 3rd edn (Freeman, San Francisco, 1995).
- Glancey, B. M., Vandenburgh, M. K. & St Romain, M. K. Testis degeneration in the imported fire ant, *Solenopsis invicta*. *J. Georgia Entomol. Soc.* **2**, 83–88 (1976).
- Keller, L. & Ross, K. G. Phenotypic plasticity and cultural transmission of alternative reproductive strategies in the fire ant *Solenopsis invicta*. *Behav. Ecol. Sociobiol.* **33**, 121–129 (1993).

**Acknowledgements.** We thank A. Bourke, C. DeHeer, J. Evans, M. Goodisman, D. Haig, L. Hurst and D. Queller for comments on the manuscript. This work was funded by grants from the Swiss and US National Science Foundations and the National Geographic Society.

Correspondence and requests for materials should be addressed to L.K. (e-mail: Laurent.Keller@izea.unil.ch).

## Visual search has no memory

Todd S. Horowitz & Jeremy M. Wolfe

Brigham & Women's Hospital and Harvard Medical School, Center for Ophthalmic Research, 221 Longwood Avenue, Boston, Massachusetts 02115, USA

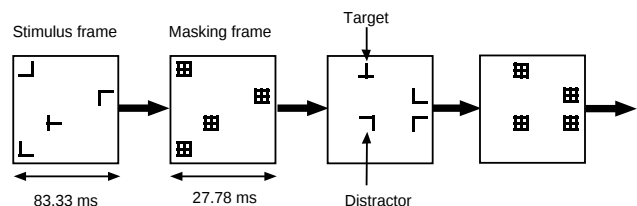
Humans spend a lot of time searching for things, such as roadside traffic signs<sup>1</sup>, soccer balls<sup>2</sup> or tumours in mammograms<sup>3</sup>. These tasks involve the deployment of attention from one item in the visual field to the next. Common sense suggests that rejected items should be noted in some fashion so that effort is not expended in re-examining items that have been attended to and rejected. However, common sense is wrong. Here we asked human observers to search for a letter 'T' among letters 'L'. This search demands visual attention and normally proceeds at a rate of 20–30 milliseconds per item<sup>4</sup>. In the critical condition, we ran-

domly relocated all letters every 111 milliseconds. This made it impossible for the subjects to keep track of the progress of the search. Nevertheless, the efficiency of the search was unchanged. Theories of visual search all assume that search relies on accumulating information about the identity of objects over time<sup>5–7</sup>. Such theories predict that search efficiency will be drastically reduced if the scene is continually shuffled while the observer is trying to search through it. As we show that efficiency is not impaired, the standard theories must be revised.

When a target item differs from distractors on a simple visual feature, such as a red bar among green bars, the target automatically grabs one's attention and can be detected independently of the number of distractor items present. When targets and distractors differ only in their spatial arrangement, however, the search becomes attention-demanding and the reaction time increases by 20–30 ms per item. Theories of visual search explain this phenomenon in one of two ways. 'Serial' models propose that attention can process the identity of only one item at a time. Once an item has been identified and rejected as a distractor, an inhibitory 'tagging' mechanism prevents that item from being revisited. As a result, a successful search for a target will require subjects to examine, on average, only half of the items in the display<sup>5</sup>. 'Parallel' theories assume that identity is computed in parallel for each item, and that an item's identity becomes gradually more certain over the course of a trial. A response is issued either when sufficient information confirms one item as the target, or when all of the items have proven to be distractors<sup>6</sup>. Both theories have in common the assumption that efficient search is based on accumulating information about the contents of the scene over the course of the trial; we refer to this as memory-driven search. We propose an alternative, that visual search processes are amnesic: they act on neural representations that are continually rewritten and have no permanent existence beyond the time span of visual persistence.

To test the hypothesis that visual search relies on memory-driven mechanisms, we designed our stimuli so that, during a trial, the scene would be constantly changing, yet the meaning of the scene (as defined by the required response) would remain constant. The task was to report as quickly as possible whether or not the target letter, T, was present in the display. In order to measure the increase in reaction time when extra items were present in the display, we varied the number of letters in the display (the set-size) between 8, 12 and 16. The slope of the target-present reaction-time × set-size function measures the efficiency of search through the display. This slope represents the added cost of each additional item. We focused on target-present slopes because their interpretation is more straightforward: the question of when to stop searching when you have not yet found a target is more complicated than the question of when to respond once you have found a target<sup>8</sup>. In half of the trials, all the letters were Ls, and these trials demanded a 'no' response. In the remaining trials, which required a 'yes' response, one of the letters was a T. Both Ts and Ls could appear, randomly, in any of four orientations: 0°, 90°, 180° or 270° to the vertical (Fig. 1).

There were two stimulus conditions in the experiments: random



**Figure 1** Two example stimulus frames from experiment 1, each followed by its corresponding masking frame. An actual trial in experiment 1 had four stimulus frames, repeated through five cycles. In experiment 2, the masking frames were eliminated and each stimulus frame was presented for 106.7 ms.

and static. In the random conditions, the stimulus locations were changed every 111 ms (Fig. 1). For any memory-driven search mechanism, this manipulation would be disastrous. It would cripple parallel accumulation of information about the identity of a particular letter. A serial model would be unable to keep track of where the letter had been, and would be forced to resample already searched locations. For a given rate of serial sampling of items, Monte Carlo simulations show that serial sampling with replacement should result in mean reaction time  $\times$  set-size slopes that are twice as steep as those resulting from the normally assumed serial sampling without replacement.

In contrast, an amnesic search mechanism would be oblivious to the randomization manipulation. For simplicity, we will describe only the serial version of an amnesic system, although an equivalent parallel interpretation can also be developed. Our model assumes that the visual system generates a priority ranking of each item in the field according to the salience of the item. This neural representation is somewhat noisy and fluctuates dynamically. Under normal circumstances, this priority ranking would reflect important feature differences in the scene (such as colour and size) and thus allow attention to be efficiently guided to the most likely target locations, effectively gating out stimuli that are unlikely to be targets<sup>9</sup>. The stimuli in these experiments were specifically designed so as not to allow such guidance.

As the representation is assumed to be noisy, there will be

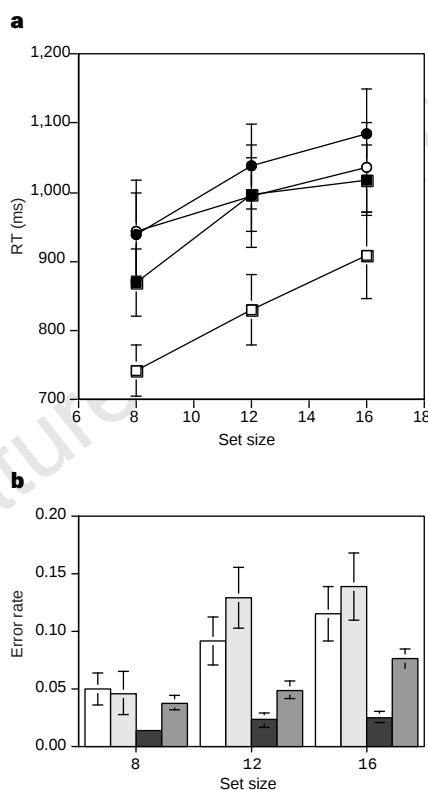
spurious differences in salience even between identical items. As a consequence, the priority assigned to each item will change over time, even in the static case. An amnesic search would proceed by determining the most likely (salient) target at the moment and directing attention to that location. If that item were to be identified as a non-target, the next item would be selected according to the same criterion, its momentary salience. From the point of view of an amnesic mechanism, there are  $n$  items at any given moment. Thus, with every sample, the system would have a  $1/n$  chance of picking out the target. When the stimuli are shuffled about between samples, a memory-driven model cannot keep track of where it has been and loses information. An amnesic mechanism, in contrast, does not keep track of items under static conditions and therefore does not lose anything when the stimuli are shuffled.

Figure 2 shows the results of correct target-present trials in two experiments; in the first experiment masked stimuli were used, in the second unmasked stimuli were used (see Methods). Subjects are slower and slightly less accurate in the random conditions. However, the slopes of the random and static target-present reaction-time  $\times$  set-size functions (Table 1) are statistically indistinguishable for both experiments ( $t_{\text{masked}}(8) = 0.13$ ,  $P < 0.50$ ;  $t_{\text{unmasked}}(8) = 1.52$ ,  $P > 0.15$ ). This agrees with the predictions of the amnesic model. In contrast to the predictions of the memory-driven-search theories, there is no evidence that subjects are searching half as efficiently in the random conditions as in the static conditions. In fact, shallower slopes are produced under random conditions than under static conditions. Data for the masked and unmasked conditions are comparable, indicating that the flickering masks in experiment 1 did not noticeably affect the search.

Although the slopes of the reaction-time  $\times$  set-size functions are no steeper in the random conditions, the mean reaction times do appear to be longer; however, the reaction-time cost is reliable in only experiment 2 ( $F(1, 8) = 18.81$ ,  $P < 0.005$ ). We suspect that any increased mean reaction times reflect subjects' decreased confidence in their responses. Consider a subject who believes she has found a target. In the static case, the physical stimulus is still available for confirmation, whereas in the random case it is not. The slope data show that, contrary to the predictions of any memory-driven account, search efficiency is similar under random and static conditions. The mean reaction time data merely indicate that subjects may be less confident in the random condition.

Error rates are higher under random conditions. This is not surprising because the random conditions are more difficult than the static conditions. However, it does raise the possibility of a speed-accuracy trade-off. Artificially shallow slopes might occur if subjects guess early in a trial rather than waiting to confirm the presence of a target. Half of the time their guesses will be correct and will contribute to a shallow slope. Half of the time they will be wrong, producing 'false alarms'. Conversely, when the target is difficult to find, subjects may give up and inaccurately respond 'no', thus unfairly taking long reaction times out of distribution.

In a third experiment, we eliminated the option to respond 'no' by having subjects respond to target identity, rather than target presence. A target letter 'E' or 'N' was present in each trial, embedded in distractors selected from the remaining letters of the alphabet (except 'I' and 'J'). Subjects identified the target letter. Otherwise, the procedure was identical to that of the unmasked experiment.



**Figure 2** Results of experiments 1 and 2. **a**, Mean correct target-present reaction times (RTs) plotted against set size for the random and static conditions from experiments 1 and 2. Squares denote the static condition and circles the random condition. Filled symbols represent experiment 1 (masked) and open symbols experiment 2 (unmasked). Error bars indicate the s.e.m. The main finding is that changing the location of items every 111 ms (random condition) does not alter the efficiency of visual search (the slope of the lines). **b**, Error rates by set size, stimulus condition and experiment. For each set size, bars from right to left indicate the masked static condition, the masked random condition, the unmasked static condition, and the unmasked random condition. Error bars indicate the s.e.m. More errors are committed under the more difficult, random conditions. However, subjects are not trading off accuracy for speed.

**Table 1** Reaction-time  $\times$  set-size slopes for experiments 1 and 2

| Target Condition      | Present Static | Present Random | Absent Static | Absent Random |
|-----------------------|----------------|----------------|---------------|---------------|
|                       | Static         | Random         | Static        | Random        |
| Masked stimuli        | 18.76 (3.66)   | 18.13 (4.66)   | 50.42 (4.25)  | 23.74 (6.75)  |
| Unmasked stimuli      | 20.89 (3.59)   | 11.51 (3.54)   | 42.00 (6.07)  | 12.18 (4.19)  |
| Target identification | 34.67 (3.20)   | 29.53 (3.02)   | NA            | NA            |

Data are shown as ms per item, means  $\pm$  s.e.m. NA, not applicable.

Once again, the slopes (29.53 ms per item in the random conditions and 34.67 ms per item in the static conditions) were statistically indistinguishable. As expected, errors were substantially lower in this experiment (5.6% errors overall for the random condition and 2.8% overall for the static) than in previous experiments. There were still more errors under random conditions than under static conditions, but not enough to make the memory-driven story plausible. For instance, if we take the subjects with the smallest differences between random and static error rates, we still find that the random and static slopes are essentially the same. Space constraints preclude a complete discussion of the speed-accuracy trade-off issue. For more details, see <http://www.dahlen.com/kari/wolfe2.html>.

Our results show that the visual system does not accumulate information about object identity over time during a search episode. Instead, the visual system seems to exist in a sort of eternal present. There is little integration of visual information across saccadic eye movements<sup>10</sup>, and observers are remarkably oblivious to dramatic scene changes when the moment of change is obscured by a brief flicker<sup>11</sup> or an intervening object<sup>12</sup>. Although subjects in such a 'change blindness' experiment may suffer momentary embarrassment, an amnesic visual system may be a handicap only in the laboratory. The structure of the world makes it unnecessary to build fully elaborated visual representations in the head. If an observer knows that an elephant was present a moment before, she can be quite sure that it will be an elephant when it is attended to again. She can guide attention to the elephant on the basis of basic features such as colour, and higher-level mnemonic strategies ('I already looked for elephants near the trees') can prevent attention from retracing its steps too often. Amnesia may be an efficient strategy for a visual system. □

## Methods

Two conditions, static and random, were tested in experiments 1–3. In experiment 1, four stimulus frames were generated for each trial (Fig. 1). If there was a target present in that trial, it was present in all frames. The same number of items was present in each frame of a given trial. In the random condition, the locations and orientations of the targets and distractors were independent from frame to frame. In the static condition, the four frames were identical; items therefore remained fixed and unchanging in their locations, alternating with masks at that location.

For each frame, we generated a corresponding masking frame by placing a mask (a square bisected along both axes) at every location at which there was a letter in the stimulus frame, such that all possible line segments of the letters would be masked. Each frame was presented for 83.33 ms, and was followed for 27.78 ms by its masking frame (Fig. 1). The cycle of four frames was repeated every 444.44 ms, and subjects observed 5.25 cycles (21 stimulus frames). Subjects were instructed to respond as soon as possible whether or not the target was present, but were allowed to respond at any time within 5,000 ms of stimulus onset. The final frame was always a mask. Subjects were allowed to respond after the final frame, but such responses accounted for less than 2% of target-present responses. Nine subjects were tested for 200 trials in each condition, randomly distributed over three set sizes (8, 12 or 16 items per display).

The second experiment differed in two respects. First, 20 frames were generated for each trial, and no frame was presented more than once, so the total presentation time was shortened to 2,140 ms. Second, we eliminated the masking frames, so each stimulus frame was now shown for 107 ms total. Stimulus exposure time was thus roughly equal across the two experiments. Nine subjects were tested for 480 trials in each condition, using the same set sizes as in experiment 1.

Both experiments were designed to thwart a strategy of attending at one location, waiting for the target to appear in the random case. In experiment 1, the target only appeared at four locations, so such a strategy would lead to failure in 93.75% of trials. In experiment 2, the target changed location in every trial but remained at the same eccentricity, so a 'sit and wait' strategy would still fail in 75% of trials.

In experiment 3, subjects identified which of two targets, an E or an N, was presented in each trial. Eleven subjects participated for 480 trials in each condition. Methods were otherwise similar to those used for experiment 2.

Received 3 April; accepted 22 May 1998.

1. Fairclough, S. & Maternaghen, M. in *Visual Search 2* (eds Brogan, D., Gale, A. & Carr, K.) 419–431 (Taylor & Francis, London, 1993).
2. Helsen, W. & Pauwels, J. M. in *Visual Search 2* (eds Brogan, D., Gale, A. & Carr, K.) 379–388 (Taylor & Francis, London, 1993).
3. Gale, A. G. & Walker, G. E. in *Visual Search 2* (eds Brogan, D., Gale, A. & Carr, K.) 231–238 (Taylor & Francis, London, 1993).
4. Wolfe, J. M. What can 1 million trials tell us about visual search? *Psychol. Sci.* **9**, 33–39 (1998).
5. Treisman, A. & Gelade, G. A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136 (1980).
6. Palmer, J. Attention in visual search: distinguishing four causes of a set-size effect. *Curr. Dir. Psychol. Sci.* **4**, 118–123 (1995).
7. Grossberg, S., Mingolla, E. & Ross, W. D. A neural theory of attentive visual search: interactions of boundary, surface, spatial, and object representations. *Psychol. Rev.* **101**, 470–489 (1994).
8. Chun, M. M. & Wolfe, J. M. Just say no: how are visual searches terminated when there is no target present? *Cogn. Psychol.* **30**, 39–78 (1996).
9. Wolfe, J. M. Guided search 2.0: a revised model of visual search. *Psychonomic Bull. Rev.* **1**, 202–238 (1994).
10. Irwin, D. E. Integrating information across saccadic eye movements. *Curr. Dir. Psychol. Sci.* **5**, 94–100 (1996).
11. Rensink, R. A., O'Regan, J. K. & Clark, J. J. To see or not to see: the need for attention to perceive changes in scenes. *Psychol. Sci.* **8**, 368–373 (1997).
12. Simons, D. J. & Levin, D. T. Change blindness. *Trends Cogn. Sci.* **1**, 261–267 (1997).

**Acknowledgements.** We thank L. Adams, K. Dahlen, H. Egeth, G. Gancarz, I. Horowitz, J. Mangels, R. Klein, A. Treisman, R. Ward and S. Yantis for comments on the manuscript, and N. Klemper for assistance with data collection. Supported by grants from AFOSR, NIH and NSF (to J.M.W.) and an NIH individual NRSA (T.S.H.).

Correspondence and requests for materials should be addressed to T.S.H. (e-mail: toddh@search.bwh.harvard.edu).

## Facilitation of long-term potentiation and memory in mice lacking nociceptin receptors

Toshiya Manabe\*†, Yukihiro Noda†‡, Takayoshi Mamiya†, Hiroyuki Katagiri\*, Takeshi Houtani‡, Miyuki Nishi§, Tetsuo Noda||, Tomoyuki Takahashi\*, Tetsuo Sugimoto‡, Toshitaka Nabeshima† & Hiroshi Takeshima§

\*Department of Neurophysiology and †Department of Pharmacology, Faculty of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113-8654, Japan

‡Department of Neuropsychopharmacology and Hospital Pharmacy, Nagoya University, School of Medicine, Nagoya 466-8560, Japan

§Department of Anatomy, Kansai Medical University, Moriguchi, Osaka 570-8506, Japan

||Department of Cell Biology, Cancer Institute, Kami-Ikebukuro, Toshima-ku, Tokyo 170-8455, Japan

¶ These authors contributed equally to this work.

The peptide nociceptin (also named orphanin FQ) acts in the brain to produce various pharmacological effects, including hyperalgesia and hypolocomotion<sup>1,2</sup>. The nociceptin receptor uses guanine-nucleotide-binding proteins to mediate the inhibition of adenylyl cyclase, the activation of potassium channels and inhibition of calcium channels<sup>3</sup>. It has been shown using knock-out mice that the nociceptin receptor is not required for regulation of nociceptive responses or locomotion activity, but modulates the auditory function<sup>4</sup>. Here we show that mice lacking the nociceptin receptor possess greater learning ability and have better memory than control mice. Histological analysis revealed the expression of both the nociceptin precursor and the nociceptin receptor in the hippocampus, thought to take part in aspects of learning and memory. Moreover, the receptor-deficient mice showed larger long-term potentiation in the hippocampal CA1 region than control mice, without apparent changes in presynaptic or postsynaptic electrophysiological properties. These results

show that the loss of the nociceptin receptor results in a gain-of-function mutation in both the memory process and the long-term potentiation mechanism in CA1, perhaps as a result of altered intracellular signal transduction systems in neurons.

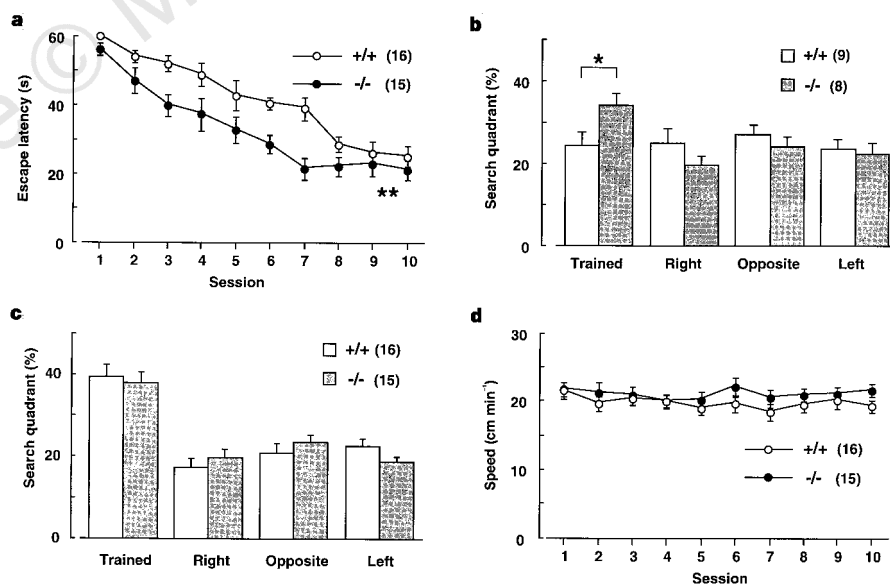
Mutant mice lacking the nociceptin receptor have the mixed genetic background of C57BL and 129 strains, are healthy and fertile, and show no significant anatomical defects<sup>4</sup>. To search for abnormal spatial learning and memory, we tested the mutant mice in the frequently used Morris water maze<sup>5,6</sup>. Both mutant and wild-type mice managed to learn the hidden-platform task, but surprisingly we found that mutants took a shorter time to reach the platform than control mice (Fig. 1a). Furthermore, in the transfer test after four days of training, the mutants searched preferentially in the trained quadrant but controls did not (Fig. 1b). This increased ability of the mutants reflected neither hyperlocomotion nor improved swimming ability because the speed of movement in the hidden-platform task was similar for mutants and wild types (Fig. 1d) and no difference was observed in the visible-platform task (data not shown). However, the test after ten days of training showed that both mutants and controls explored the correct position predominantly, no significant difference being observed (Fig. 1c). These results suggest that the knockout mice learned the water maze task faster than the wild-type mice, but that there are no differences in the reachable capacity or the retention of the spatial memory between the two groups.

To examine further the ability of the nociceptin receptor-deficient mice, we tested them in the step-through type passive avoidance task, an index of learning and memory of distasteful experience<sup>7</sup>. In this task, a mouse was placed in a cage consisting of light and dark compartments, and received an electric shock when it entered the dark chamber to make it stay in the light chamber for a constant time during the training trials. After several days, the step-through latencies to the dark chamber were recorded in the retention trials. In the training trials, we found no difference between the knockout

and wild-type mice either in the step-through latency on first training (Fig. 2a) or in the amount of training required to learn the task (Fig. 2b). However, the step-through latencies of the mutants were significantly longer than those of controls in the retention trials (Fig. 2c). It is unlikely that the prolonged latency of the mutants is due to the heterogeneity in genetic background because C57BL and 129/sv mice showed essentially similar responses in this task (data not shown). Therefore, the mutant mice exhibited increased retention of the passive avoidance memory.

It is known that spatial learning and memory require integrative control functions of the hippocampus<sup>8</sup>. Nissl-stained material from the entire hippocampal region of the nociceptin receptor-deficient mice showed no histological abnormality. Its  $\beta$ -galactosidase staining showed that the reporter message for the nociceptin receptor in the mutant mice<sup>4</sup> is highly expressed in the dentate gyrus granule cells and in CA1 and CA3 pyramidal cells (Fig. 3a). *In situ* hybridization analysis showed that nociceptin precursor is produced in the hippocampal cells (Fig. 3b). These results suggest that both CA1 and CA3 pyramidal cells are normally enriched with nociceptin and its receptor, that the interaction between nociceptin and receptor may occur readily in these cells or with the supply of the ligand by nearby pyramidal cells, and that the nociceptin system may also enable the regulation of CA1 neurons by CA3 neurons.

To examine synaptic plasticity in the nociceptin receptor-deficient mice, we analysed long-term potentiation (LTP) in the CA1 region using hippocampal slices (Fig. 4a, b). High-frequency stimulation caused stable potentiation of excitatory postsynaptic potentials (EPSPs) in wild-type mice ( $133.5 \pm 7.8\%$ ), whereas the same conditioning gave rise to significantly larger LTP in the mutant mice ( $158.4 \pm 8.5\%$ ;  $P = 0.021$ , Student's *t*-test). These results indicate that the nociceptin receptor plays a role in the negative regulation of CA1 LTP induction and that the increased LTP caused by lack of receptor is not compensated by other modulation



**Figure 1** Performance in Morris water-maze task in nociceptin receptor-deficient and wild-type mice. **a**, Escape latency in the hidden-platform test. The mutant mice ( $-/-$ ) showed significantly shorter escape latencies than control mice ( $+/+$ ) on the sessions ( $**F(1,29) = 20.087$ ,  $P = 0.001$ , ANOVA). The difference of the escape latency on the first day between the two groups is due to the tendency of the mutant mice to reach the platform faster than control mice in the third trial. **b**, Transfer test performed one day after the four-day training period. In the mutant mice, the time spent in the trained quadrant was significantly longer than that in any other quadrant ( $P < 0.05$ , Dunnett's multiple comparison test). The mutant mice spent significantly longer time in the trained quadrant than control mice

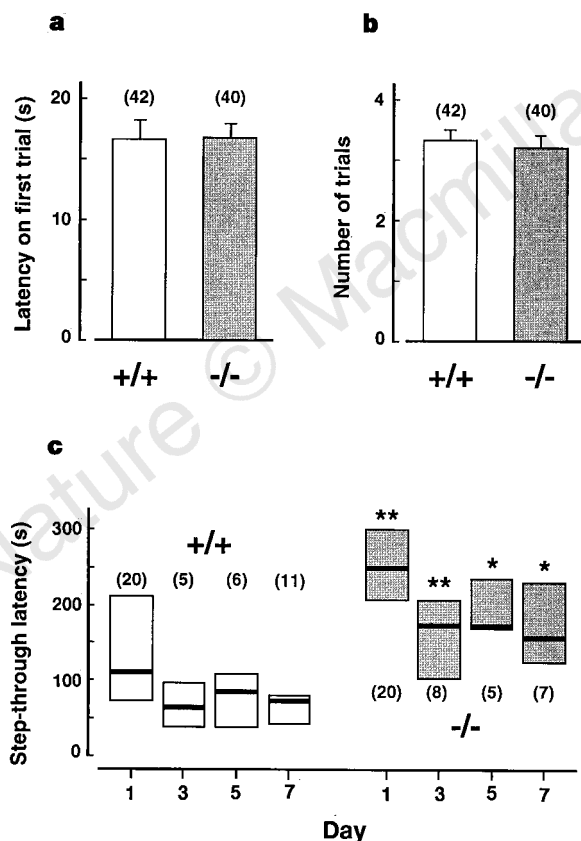
( $*P = 0.028$ , Student's *t*-test). **c**, Transfer test performed three days after the ten-day training period. Both mutant and control mice took significantly longer time in the trained quadrant than in any other quadrant (in both groups  $P < 0.01$ , Dunnett's multiple comparison test), but there were no significant differences between the genotypes. **d**, Speed of swimming in the hidden-platform test. Moving distances for 60 s were calculated from recorded swimming paths. No significant difference was observed between the mutant and wild-type mice. Data represent mean  $\pm$  s.e.m. and the numbers of tested animals are indicated in parentheses.



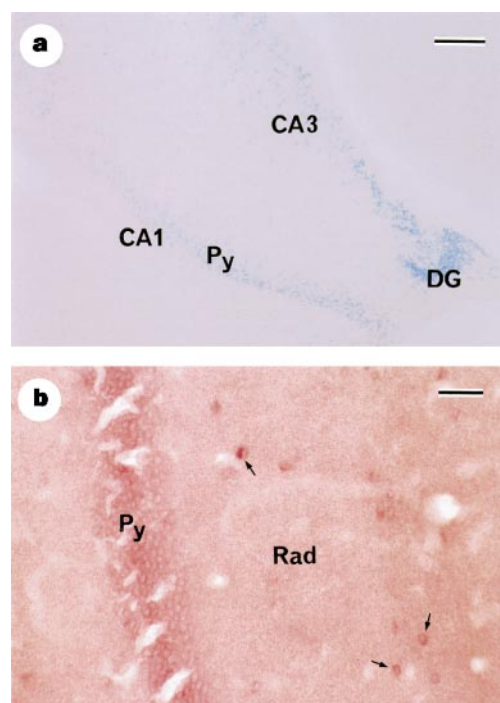
systems. Activation of nociceptin receptors in hippocampal CA3 and CA1 neurons is reported to inhibit  $\text{Ca}^{2+}$  channels, including N and P/Q types<sup>9</sup>, that mediate transmitter release from the hippocampal presynaptic terminal<sup>10</sup>. Thus greater LTP in the mutants could be due to facilitated release of glutamate during high-frequency stimulation of the presynaptic fibres, which would activate glutamate receptors required for LTP induction more efficiently. However, paired-pulse facilitation induced by a pair of stimulations of afferent fibres at a short interval (50 ms), which is sensitive to presynaptic release probability<sup>11</sup>, was indistinguishable between wild-type ( $1.36 \pm 0.07$ ,  $n = 7$  slices from 5 mice) and mutant mice ( $1.43 \pm 0.05$ ,  $n = 10$  slices from 6 mice). Another presynaptic short-term plasticity or post-tetanic potentiation following tetanic stimulation applied in the presence of D-2-amino-5-phosphonvaleric acid (AP5), a competitive NMDA (N-methyl-D-aspartate) receptor antagonist, in mutants was similar to that in controls (Fig. 4c). Thus, it is unlikely that the nociceptin receptor modulates LTP through presynaptic mechanisms. However, recent findings demonstrating functional coupling of nociceptin receptors with inwardly rectifying  $\text{K}^+$  channels in CA1 and CA3 pyramidal cells<sup>12</sup> suggest that the lack of the nociceptin receptor may cause larger depolarization during tetanic stimulation; this would increase the activation of NMDA receptors by removing

the  $\text{Mg}^{2+}$  block<sup>13,14</sup>. However, the envelope of depolarization during tetanic stimulation was not significantly different between mutant and wild-type mice (Fig. 4d). Moreover, there was no difference between the two groups in the resting membrane potential of CA1 pyramidal cells measured with whole-cell recording techniques (T. Manabe and H.K., unpublished observations). These results suggest that changes in the postsynaptic electrophysiological parameters are not associated with increased LTP in the mutant mice.

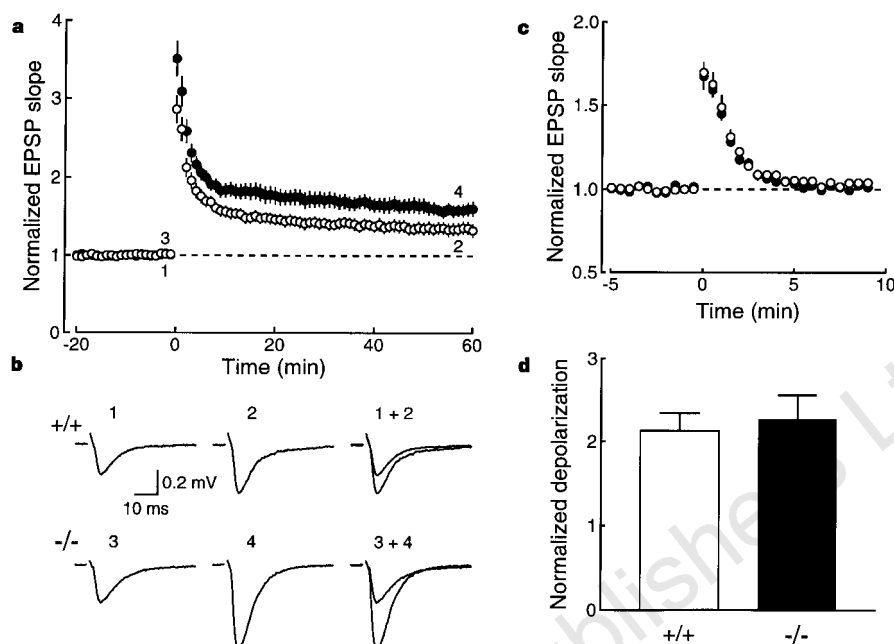
We recently found that nociceptin receptor-deficient mice outperformed wild-type mice in the water-finding task<sup>15</sup>, an index of spatial attention and latent memory<sup>16</sup>. Taking into account our latest results, three independent behavioural tasks indicated the facilitation of learning and memory processes in the mutants, and the conclusion may be supported by the data that nociceptin injected into the hippocampal region results in impaired spatial learning in rats<sup>17</sup>. However, our experimental data seem to be slightly contradictory: the water-maze test suggests that mutants have an improved memory-acquirement process, whereas the passive avoidance and water-finding tests indicate an improved memory-retention process. Difference in motivation for the tasks may be underlying this contradiction, and these tasks may not distinguish adequately between learning and memory processes. The data suggesting that the application of nociceptin inhibits CA1 LTP<sup>18</sup> may support our studies demonstrating increased LTP in the mutants. However, our results also suggest that there are no significant abnormalities in presynaptic or postsynaptic electrophysiological properties in the mutants.



**Figure 2** Performance in multiple-trial passive avoidance test in nociceptin receptor-deficient and wild-type mice. **a**, Step-through latency of mutant (–/–) and wild-type (+/+) mice on the first trial in the acquisition trials. **b**, Number of trials in the acquisition trials. There were no significant differences between the two groups of mice in the step-through latency on the first trial and the number of trials required to reach the criterion of staying in the light compartment for 120 s. **c**, Step-through latencies in the retention trials performed 1, 3, 5 and 7 days after the acquisition trials. The step-through latencies of the mutant mice were significantly prolonged compared with those of wild-type mice (\*\* $P < 0.01$  and \* $P < 0.05$ , Student's *t*-test). Each value represents the mean  $\pm$  s.e.m. in **a** and **b**, or the median (bold line) and interquartile ranges (column) in **c**. Numbers of tested animals are indicated in parentheses.



**Figure 3** Expression of nociceptin precursor and nociceptin receptor in hippocampal regions. **a**, Expression of  $\beta$ -galactosidase activity as the reporter message for the nociceptin receptor in the nociceptin receptor-deficient mice. A high level of the enzyme activity is seen throughout the granule cells of the dentate gyrus (DG) and in the CA1 and CA3 pyramidal cells (Py). A similar staining pattern was found in the heterozygous mice; the wild-type mice lacked enzyme activity. **b**, Expression of nociceptin precursor mRNA in the mouse hippocampal CA1 region. The hybridization signal is notable in many pyramidal cells (Py) and scattered neurons in the stratum radiatum (Rad, arrows). Similar positive staining was obtained for the CA3 region; all the hippocampal regions failed to be stained with a sense probe. Scale bars: 200  $\mu\text{m}$  in **a**, 50  $\mu\text{m}$  in **b**.



**Figure 4** Enhanced LTP in nociceptin receptor-deficient mice. **a**, The averaged time course of LTP in mutant (filled circles;  $n = 14$  slices from 6 mice) and wild-type (open circles;  $n = 16$  slices from 8 mice) mice. Initial EPSP slopes were normalized in each experiment using the averaged slope value during the control period (-20 to 0 min). The potentiation ratio was calculated using the slope value from 50 to 60 min. Tetanic stimulation (100 Hz for 1 s) was applied at time 0. **b**, Sample traces of field EPSPs (average of 10 consecutive responses) of the mutant (-/-) and wild-type (+/+) mice recorded at the times indicated in **a**. **c**, Post-

tetanic potentiation measured in the presence of 50  $\mu$ M AP5 in wild-type (open circles,  $n = 7$  slices from 4 mice) and mutant (filled circles;  $n = 10$  slices from 5 mice) slices. Tetanic stimulation was delivered at time 0. **d**, Depolarization during tetanic stimulation was not different between wild-type ( $n = 7$  slices from 5 mice) and mutant ( $n = 10$  slices from 6 mice) slices. Depolarization at 200 ms from the beginning of the tetanus was normalized among slices by dividing it by the amplitude of baseline EPSPs before tetanic stimulation (normalized depolarization,  $2.13 \pm 0.21$  in wild-type mice and  $2.27 \pm 0.29$  in mutant mice).

It has been reported previously that the nociceptin receptor mediates the downregulation of cAMP formation<sup>1,2</sup> and that cAMP increases excitability by inhibiting  $\text{Ca}^{2+}$ -activated  $\text{K}^{+}$  channels<sup>19</sup> and facilitates CA1 LTP induction<sup>20–22</sup>. Nociceptin released by high-frequency stimulation may modulate LTP induction through the regulation of postsynaptic cAMP levels, and the lack of negative regulation in cAMP levels may cause the LTP enhancement in the mutants. Alternatively, abnormalities of the signal transduction systems involving protein kinase C or mitogen-activated protein kinase might cause increased LTP in the mutants, because previous studies suggest that such protein kinases modulate CA1 LTP postsynaptically<sup>23,24</sup> and are activated by stimulation of the nociceptin receptor<sup>25,26</sup>.

In conclusion, the nociceptin system seems to play negative roles both in learning and memory at the whole-animal level and in CA1 LTP at the cellular level. Our studies suggest that memory processes can be modulated by ligands to the nociceptin receptor, and further that the antagonists are worth testing for the alleviation of memory disorders. □

## Methods

**Mice.** Nociceptin receptor-deficient mice carrying the *mor1*<sup>ml</sup>/*mor1*<sup>ml</sup> genotype (6–12 weeks old) were obtained by crossing between the heterozygotes<sup>4</sup>, and wild-type littermates were used as controls for the behavioural studies.

**Water-maze tests.** For the Morris water-maze task<sup>5,6</sup>, a pool (1.2 m in diameter) was prepared with white polypropylene plastics and water maintained at 18 °C. In the hidden-platform test, the Plexiglas platform (7 cm in diameter) was submerged 2 cm below water level and kept in the same location throughout training. Swimming paths were tracked with a camera fixed on the ceiling of the room and stored in a computer (Target/2 system, Neuroscience Inc). The mice did not swim in the pool before the training. Three starting positions were used pseudo-randomly and the mice were trained with three

trials per day for four or ten days. After reaching the platform, the mouse was allowed to remain on it for 30 s before being returned to its home cage. If the mouse did not find the platform within 60 s, the training trial was terminated and a maximum score of 60 s was assigned. In a transfer test, mice swam for 60 s in the pool without the platform. In the visible-platform test, the black platform (7 cm in diameter) was located 1 cm below water level.

**Passive avoidance tests.** For the step-through type passive avoidance test<sup>7</sup>, mice were examined in the apparatus, which consisted of illuminated and darkened compartments (16 × 11 cm base, 11 cm high) with a grid floor and a guillotine door separating the compartments. In the acquisition trial, each mouse was placed in the light compartment and allowed to explore it for 10 s. The door to the dark compartment was opened, and the time before the mouse entered the dark compartment (step-through latency) was recorded. When the mouse stepped through the door, the door was closed and an electric shock (0.2 mA, 2 s, Shockgenerator-Scrambler SG-NS01, Neuroscience Inc) was delivered through the grid. This training continued until the mouse stayed in the light compartment for 120 s. The retention trial was carried out several days after the training and the step-through latency was recorded for up to 300 s. The number of training trials in the acquisition trials and the step-through latency in the retention trial were used as the index of acquisition and retention of the training experience, respectively. We found no difference in the behavioural responses to varied electric shocks between the nociceptin receptor-deficient and wild-type mice<sup>15</sup>.

**Histological studies.** For  $\beta$ -galactosidase staining, mice (8 weeks old) were anaesthetized deeply and perfused with a 2% paraformaldehyde solution buffered with 0.12 M sodium phosphate (pH 7.4). The brains were immersed immediately in ice-cold 25% sucrose. Frozen coronal sections (40  $\mu$ m thickness) were stained with the X-Gal reagent<sup>27</sup>. For *in situ* hybridization analysis, mice were perfused with a fixative containing 4% paraformaldehyde, 2% glutaraldehyde and 0.12 M sodium phosphate (pH 7.4) under pentobarbital anaesthesia. Coronal brain sections (50  $\mu$ m thick) were subjected to *in situ* hybridization using the digoxigenin-labelled riboprobe for nociceptin precursor mRNA<sup>28</sup>.

**Electrophysiological studies.** To compare the mutant and wild-type mice, most experiments were performed in a blind fashion, and the results were essentially identical to those of non-blind experiments so all data were pooled. Hippocampal slices (400  $\mu\text{m}$  thick) prepared from mice 5–9 weeks old were placed in a holding chamber for at least 1 h. A single slice was then transferred to the recording chamber and submerged beneath a continuously perfusing medium that had been saturated with 95%  $\text{O}_2$ /5%  $\text{CO}_2$ . The medium contained 119 mM NaCl, 2.5 mM KCl, 1.3 mM  $\text{MgSO}_4$ , 2.5 mM  $\text{CaCl}_2$ , 1.0 mM  $\text{NaH}_2\text{PO}_4$ , 26.2 mM  $\text{NaHCO}_3$ , 11 mM glucose. All perfusing solutions contained 100  $\mu\text{M}$  picrotoxin to block GABA<sub>A</sub> receptor-mediated inhibitory synaptic responses. Field-potential recordings were made using a glass electrode (3 M NaCl) placed in the stratum radiatum. An Axopatch 1D amplifier was used and the signal was filtered at 1 kHz and digitized at 10 kHz. For evoking synaptic responses, a bipolar tungsten stimulating electrode was placed in the stratum radiatum and Schaffer collateral/commissural fibres were stimulated at 0.1 Hz. All experiments were done at room temperature (25–28 °C). The data are expressed as means  $\pm$  s.e.m. Student's *t*-tests were used to determine whether there was a significant difference ( $P < 0.05$ ) in the mean between two sets of data.

Received 22 April; accepted 19 May 1998.

- Meunier, J.-C. *et al.* Isolation and structure of the endogenous agonist of the opioid receptor-like ORL1 receptor. *Nature* **377**, 532–535 (1995).
- Reinscheid, R. K. *et al.* Orphanin FQ: a neuropeptide that activates an opioidlike G protein-coupled receptor. *Science* **270**, 792–794 (1995).
- Henderson, G. & McKnight, A. T. The orphan opioid receptor and its endogenous ligand-nociceptin/orphanin FQ. *Trends Pharmacol. Sci.* **18**, 293–300 (1997).
- Nishi, M. *et al.* Unrestrained nociceptive response and dysregulation of hearing ability in mice lacking the nociceptin/orphaninFQ receptor. *EMBO J.* **16**, 1858–1864 (1997).
- Morris, R. G. M. Spatial localization does not depend on the presence of local cues. *Learn. Motiv.* **12**, 239–260 (1981).
- Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327–1338 (1996).
- Yamada, K. *et al.* The role of nitric oxide in dizocilpine-induced impairment of spontaneous alternation behavior in mice. *J. Pharmacol. Exp. Ther.* **276**, 460–466 (1996).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Knoflach, F., Reinscheid, R. K., Civelli, O. & Kemp, J. A. Modulation of voltage-gated calcium channels by orphanin FQ in freshly dissociated hippocampal neurons. *J. Neurosci.* **16**, 6657–6664 (1996).
- Takahashi, T. & Momiyama, A. Different types of calcium channels mediate central synaptic transmission. *Nature* **366**, 156–158 (1993).
- Manabe, T., Wyllie, D. J. A., Perkel, D. J. & Nicoll, R. A. Modulation of synaptic transmission and long-term potentiation; effects on paired pulse facilitation and EPSC variance in the CA1 region of the hippocampus. *J. Neurophysiol.* **70**, 1451–1459 (1993).
- Ikeda, K. *et al.* Functional coupling of the nociceptin/orphanin FQ receptor with the G-protein-activated K<sup>+</sup> (GIRK) channel. *Mol. Brain Res.* **45**, 117–126 (1997).
- Nowak, L., Bregestovski, P., Ascher, P., Herbert, A. & Prochiantz, A. Magnesium gates glutamate-activated channels in mouse central neurons. *Nature* **307**, 462–465 (1984).
- Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. Voltage-dependent block by  $\text{Mg}^{2+}$  of NMDA responses in spinal cord neurons. *Nature* **309**, 261–263 (1984).
- Mamiya, T., Noda, Y., Nishi, M., Takeshima, H. & Nabeshima, T. Enhancement of spatial attention in nociceptin receptor-knockout mice. *Brain Res.* **783**, 236–240 (1998).
- Nabeshima, T. & Ichihara, K. *Paradigms for the Study of Behavior*, Vol. 14 (ed. Conn, P. M.) 217–229 (Academic, San Diego, 1993).
- Sandin, J., Georgieva, A. J., Schott, P. A., Ogren, S. O. & Terenius, L. Nociceptin/orphanin FQ microinjected into hippocampus impairs spatial learning in rats. *Eur. J. Neurosci.* **9**, 194–197 (1997).
- Yu, T. P., Fein, J., Phan, T., Evans, C. J. & Xie, C. W. Orphanin FQ inhibits synaptic transmission and long-term potentiation in rat hippocampus. *Hippocampus* **7**, 88–94 (1997).
- Madison, D. V. & Nicoll, R. A. Noradrenaline blocks accommodation of pyramidal cell discharge in the hippocampus. *Nature* **299**, 636–638 (1982).
- Frey, U., Huang, Y.-Y. & Kandel, E. R. Effects of cAMP simulate a late phase of LTP in hippocampal CA1 neurons. *Science* **260**, 1661–1664 (1993).
- Blitzer, R. D., Wong, T., Nouranifar, R., Iyengar, R. & Landau, E. M. Postsynaptic cAMP pathway gates early LTP in hippocampal CA1 region. *Neuron* **15**, 1403–1414 (1995).
- Thomas, M. J., Moody, T. D., Makhinson, M. & O'Dell, T. J. Activity-dependent  $\beta$ -adrenergic modulation of low frequency stimulation induced LTP in the hippocampal CA1 region. *Neuron* **17**, 475–482 (1996).
- Hu, G.-Y. *et al.* Protein kinase C injection into hippocampal pyramidal cells elicits features of long-term potentiation. *Nature* **328**, 426–429 (1987).
- Malinow, R., Schulman, H. & Tsien, R. W. Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of LTP. *Science* **245**, 862–866 (1989).
- Lou, L.-G., Ma, L. L. & Pei, G. Nociceptin/orphanin FQ activates protein kinase C, and this effect is mediated through phospholipase C/ $\text{Ca}^{2+}$  pathway. *Biochem. Biophys. Res. Commun.* **240**, 304–308 (1997).
- Fukuda, K., Shoda, T., Morikawa, H., Kato, S. & Mori, K. Activation of mitogen-activated protein kinase by the nociceptin receptor expressed in Chinese hamster ovary cells. *FEBS Lett.* **412**, 290–294 (1997).
- Bajocchi, G., Feldman, S. H., Crystal, R. G. & Mastrangeli, A. Direct *in vivo* gene transfer to ependymal cells in the central nervous system using recombinant adenovirus vector. *Nature Genet.* **3**, 229–234 (1993).
- Houtani, T., Nishi, M., Takeshima, H., Nukada, T. & Sugimoto, T. Structure and regional distribution of nociceptin/orphanin FQ precursor. *Biochem. Biophys. Res. Commun.* **219**, 714–719 (1996).

**Acknowledgements.** We thank A. Sakamoto and I. Dobashi for help maintaining the mutant mice. This work was supported in part by grants from the Ministry of Education, Science, Sports and Culture of Japan, the Japan Private School Promotion Foundation, the Uehara Memorial Foundation, the TMFC and the Life Science Foundation.

Correspondence and requests for materials should be addressed to H.T. (e-mail: takeshim@m.u-tokyo.ac.jp).

## Kinetics and regulation of fast endocytosis at hippocampal synapses

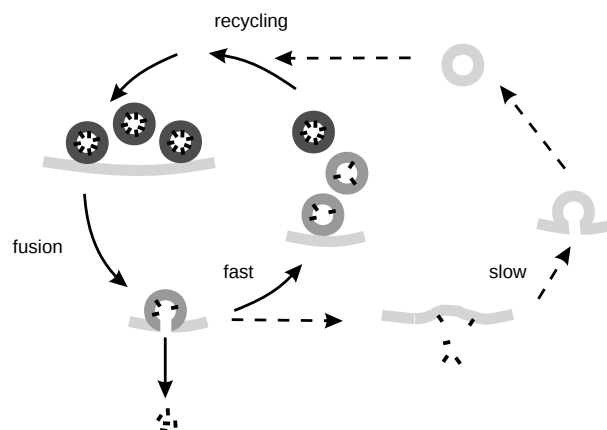
Jürgen Klingauf\*, Ege T. Kavalali\* & Richard W. Tsien

Department of Molecular and Cellular Physiology, Stanford Medical Center, Stanford, California 94305, USA

\* These authors contributed equally to this work.

Presynaptic nerve terminals often contain as few as a hundred vesicles<sup>1,2</sup> and so must recycle them soon after exocytosis to preserve synaptic transmission and presynaptic morphology<sup>3,4</sup> during repetitive firing. The kinetics<sup>3,4</sup> and mechanisms<sup>5</sup> of vesicular endocytosis and repriming have therefore been studied. Vesicles in hippocampal nerve terminals can become available to release their contents within  $\sim 40$  s of the previous round of exocytosis<sup>6,7</sup>. Studies using the styryl dye FM1-43 (ref. 3) have estimated the time constant for endocytosis as  $\sim 20$ –30 s (refs 4, 8), at least half of the total recycling time, which is much slower than endocytosis in other secretory systems<sup>9–11</sup>. It seems paradoxical that the neurosecretory terminals that could benefit the most from rapid endocytosis do not use such a mechanism. Here we demonstrate the existence of fast endocytosis in hippocampal nerve terminals and derive its kinetics from fluorescence measurements using dyes with varying rates of membrane depolarization. The rapid mode of vesicular retrieval was much faster after exposure to staurosporine or elevated extracellular calcium. Thus hippocampal synapses take advantage of efficient mechanisms for endocytosis, and their vesicular retrieval is subject to modulatory control.

Various modes of vesicle recycling described in other systems<sup>12</sup> must be considered at synapses of the central nervous system (Fig. 1). In the classical endocytotic pathway<sup>13–15</sup>, vesicles collapse completely into the surface membrane, internalize slowly ( $>20$  s) then recycle through the endosome. In a more rapid endocytotic pathway<sup>15–19</sup>, vesicles are rapidly retrieved after short-lived fusion pore openings ('kiss and run'<sup>16,17</sup> or 'cavapture'<sup>12</sup>), or may reform directly by uncoating of clathrin-coated vesicles without equilibra-



**Figure 1** A schematic depiction of recycling of vesicles marked with FM dye. A strong stimulus causes fusion of a large fraction of stained vesicles, followed by dye loss that depends on the rates of endocytosis and dye depolarization. In the classical endocytotic pathway (slow), vesicle membrane is retrieved outside the active zone in tens of seconds and should therefore not contain any significant amount of dye. If, however, membrane is retrieved more rapidly at the sites of exocytosis (fast), some fraction of dye will be trapped in freshly retrieved vesicles, which will eventually recycle back to the primed pool.

tion with endosomes<sup>20</sup>. Unlike the classical pathway, which is slow enough to allow complete dye loss, a rapid endocytotic mechanism might internalize vesicles before dye had fully partitioned and escaped. This could be detected in two ways using fluorescent probes such as FM1-43: first, the initial fluorescence loss would fall short of completion, no matter how strong the stimulus; and second, dye retained by rapid vesicular recapture would continue to mark recycling vesicles but would eventually escape in later exocytotic events, supporting delayed destaining.

We tested these predictions in cultured hippocampal neurons<sup>7,21</sup>. Presynaptic terminals (boutons) were loaded with FM1-43, washed then perfused with a 90 mM  $K^+$ /2 mM  $Ca^{2+}$  solution to stimulate exocytosis and associated endocytosis<sup>6</sup> (Fig. 2a–c). The continuous depolarization-induced  $Ca^{2+}$  entry caused a markedly biphasic decrease in fluorescence averaged across multiple boutons (Fig. 2c). An initial phase of rapid destaining gave way within seconds to a much more gradual destaining that continued for more than 1 min. Destaining profiles of single fluorescent spots helped to explain the basis of the slow phase (Fig. 2a). In any given trial, some puncta displayed a delayed burst of fluorescence loss, a 'kink' in the fluorescence trace (Fig. 2a, arrows), whereas others did not. Delayed bursts were formally identified as sharp increases in the rate of fluorescence loss,  $>3$  standard deviations above background (Fig.

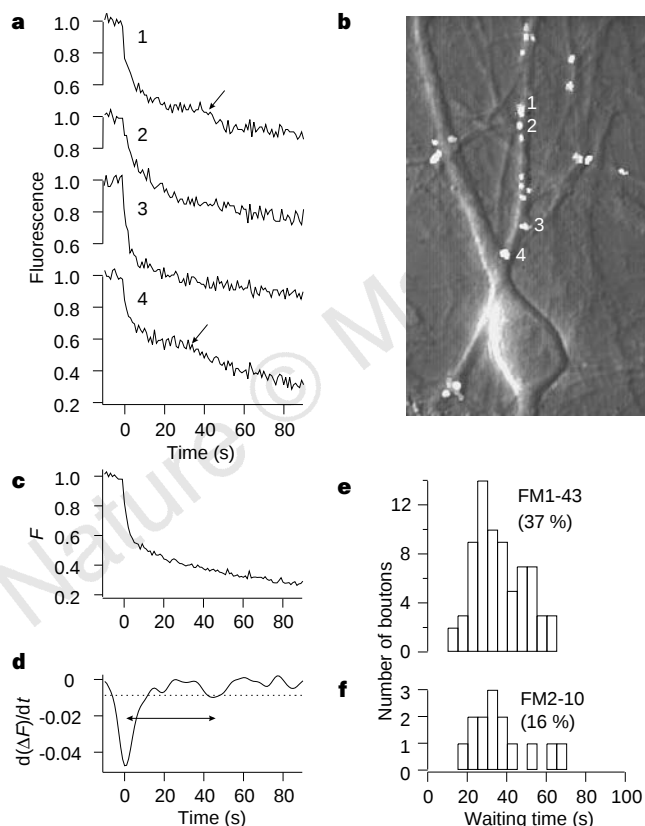
2d); these were found in 37% of the boutons and appeared 20–60 s after the initial drop in fluorescence (Fig. 2e). The mean interval agrees with the period of  $\sim 40$  s required for recycling vesicles that have just undergone exocytosis<sup>4,6,7,22</sup>. When nerve terminals were stained with FM2-10, another styryl compound with weaker affinity for membranes than FM1-43 (ref. 3), the delayed bursts appeared with similar timing, but only in 16% of the traces (Fig. 2f). The lower incidence of clear delayed bursts was as expected if lesser amounts of this dye were retained after the initial round of exocytosis.

FM2-10, FM1-43 and FM1-84 differ in terminal hydrocarbon chain length (2, 4 or 5 carbons; Fig. 3a), and exhibit different rates of membrane dissociation<sup>4</sup>. This allowed the dyes to be used to estimate the rate of rapid endocytosis. Exponential time constants of membrane departitioning, determined by destaining somato-dendritic membranes of hippocampal neurons (Fig. 3a), increased from 0.6 to 4.7 s with increasing hydrocarbon chain length. The same dyes showed significantly different destaining behaviour in hippocampal boutons subjected to continuous depolarization-induced  $Ca^{2+}$  entry (Fig. 3b). The destaining 15 s after initiating the stimulus ranged from 59% for the rapidly departitioning FM2-10 to 46% for FM1-43 and 42% for FM1-84, the slowest of the three probes. The differences in dye loss persisted for many tens of seconds, far longer than any time constant for dye departitioning, consistent with dye being trapped in internalized vesicles. This was verified by application of several stimuli in succession (Fig. 3c). The fluorescence signal hardly changed during the  $\sim 60$  s quiescence between stimuli, but dropped immediately when depolarization was resumed, as expected for a new round of  $Ca^{2+}$ -dependent exocytosis from internalized vesicles. Although destaining was larger for FM2-10 than FM1-43 during the first trial, it was greater for FM1-43 than FM2-10 during the second and third trials (Fig. 3c, inset).

Destaining was less rapid when evoked by field stimulation at 10 Hz (Fig. 3d) than with high  $K^+$ / $Ca^{2+}$  stimulation. Nonetheless, the degree of early destaining showed the same systematic variation with the dyes' departitioning kinetics. This finding reinforced the idea that a substantial fraction of synaptic vesicles undergo endocytosis quickly enough to retain the dye. With even milder stimulation (1 Hz field stimulation), kinetic distinctions among the three dyes were no longer significant<sup>4</sup> (Fig. 3d, inset).

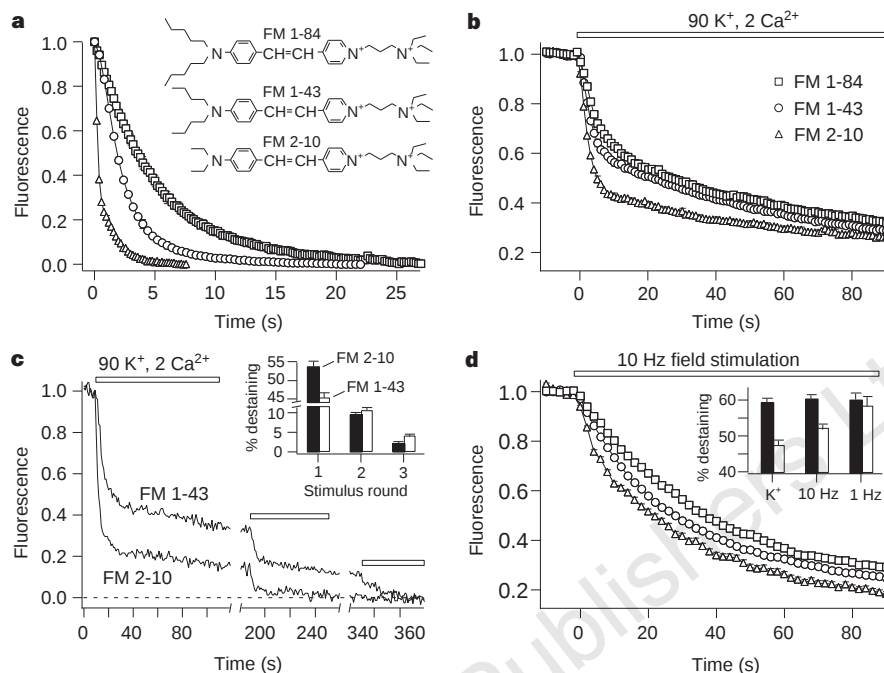
A simplified kinetic model (Fig. 4a) yielded predictions for the time course of dye destaining and sequestration that depended critically on  $\beta$ , the rate of rapid endocytosis. Reasonable fits to the experimental data were obtained (Fig. 4b) if we assumed that 60% of the vesicles can be released with the time constant  $\tau$  ( $= 1/\alpha$ ) of 2 s, and that dye reuptake proceeded with an exponential time constant of  $\sim 6$  s. The spread between the destaining profiles for FM2-10, FM1-43 and FM1-84 was predicted by setting the rate constant  $\delta$  according to the observed kinetics for dye departitioning (Fig. 3a). Simulation of a delayed burst of destaining was achieved by breaking up the transition  $\gamma$  into smaller steps with the same overall rate (Fig. 4c), allowing for a series of events in recycling of retrieved vesicles<sup>5</sup>.

If endocytosis in hippocampal terminals proceeds several-fold faster than previously thought, might it be accelerated further by modulatory mechanisms? The non-selective kinase inhibitor staurosporine eliminates FM1-43 destaining without affecting quantal acetylcholine release at neuromuscular junctions<sup>18</sup> and reduces FM1-43 loss at hippocampal terminals<sup>22</sup>. Accordingly, we looked for possible effects of staurosporine on endocytosis (Fig. 5a, b). Preincubation with staurosporine produced a striking diminution of the initial loss of FM1-43 (Fig. 5a) and FM2-10 (Fig. 5b), without any significant effect on the total amount of dye release. Both sets of data could be accounted for by increasing the endocytotic rate constant by approximately threefold (smooth curves;  $\beta^{-1}$  reduced from 6 s to  $\sim 1.7$  s in staurosporine). In contrasts simulated effects of various changes in exocytotic rate did not fit the data. Thus the



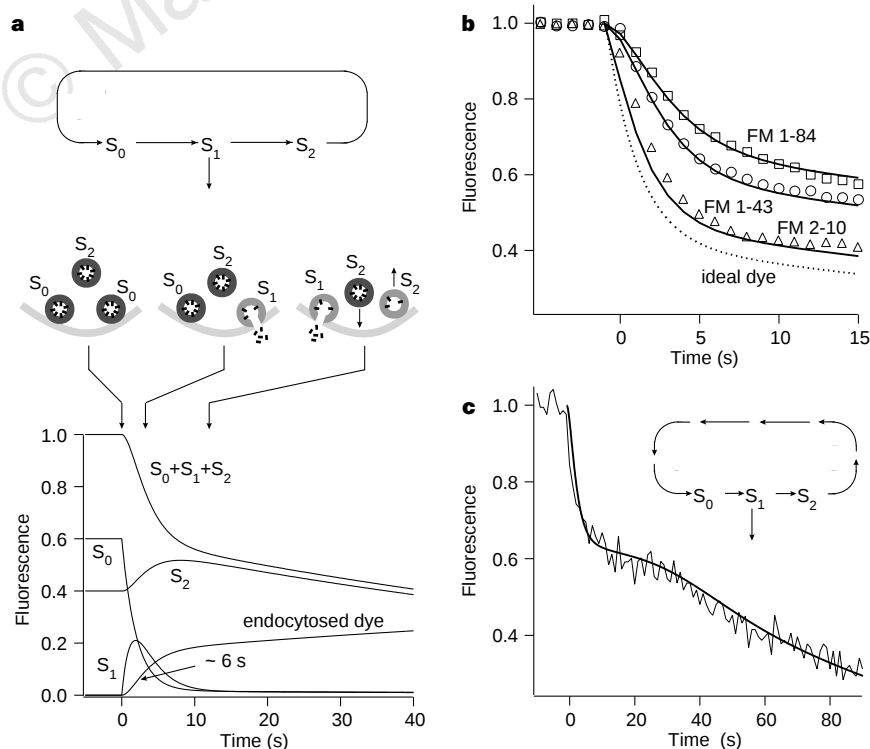
**Figure 2** Destaining characteristics of individual synaptic boutons in response to sustained 90  $K^+$ , 2  $Ca^{2+}$  stimulation. **a**, **b**, Destaining profiles from four FM1-43 loaded boutons, shown in **b**. Note secondary bursts of destaining (arrows) in traces 1 and 4. **c**, Ensemble average of destaining patterns from all boutons in **b**. The biphasic destaining (loss of fluorescence,  $\Delta F$ ) was not accounted for by  $Ca^{2+}$  channel inactivation: when  $Ca^{2+}$  channels were preinactivated by 30 s exposure to high  $K^+$ , 0  $Ca^{2+}$ , subsequent admission of  $Ca^{2+}$  still caused biphasic destaining. **d**, Secondary bursts of destaining identified in the derivative of the smoothed fluorescence signal. The broken line shows a threshold criterion for secondary bursts, three s.d. above baseline fluctuations. **e**, **f**, Waiting time distributions of secondary bursts; 72 out of 191 (38%) FM1-43-stained boutons showed secondary bursts, but only 14 out of 87 (16%) showed secondary bursts after FM2-10 staining.





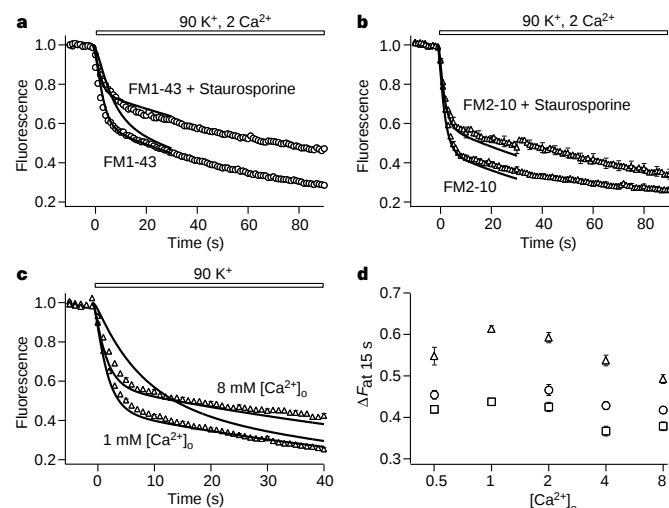
**Figure 3** Differences in departitioning kinetics of FM dyes support detection of fast endocytosis. **a**, Departitioning time constants determined by fast wash-out of equilibrated dyes from neuronal membranes (2  $\mu$ M FM1-84; 4.7 s; 4  $\mu$ M FM1-43, 2.5 s; 200  $\mu$ M FM 2-10, 0.6 s). Note correlation with alkyl chain length (inset). **b**, Ensemble averages of destaining profiles with 2  $\mu$ M FM1-84, 4  $\mu$ M FM1-43 and 200  $\mu$ M FM2-10 ( $n = 92$ ,  $n = 90$  and  $n = 87$  boutons, respectively). **c**, Smaller amplitudes of initial destaining for FM1-43 compared to FM2-10 indicate re-internalization of FM1-43. Inset shows a comparison of amplitudes of destaining, 15 s after stimulation, for multiple rounds of stimuli (first round,  $53.3 \pm 0.01\%$

versus  $46.5 \pm 0.01\%$  for FM2-10 ( $n = 90$ ) and FM1-43 ( $n = 87$ ), respectively,  $P < 0.0001$ ; second round,  $9.3 \pm 0.007\%$  versus  $10.5 \pm 0.005\%$ ,  $P = 0.07$ ; third round,  $2.1 \pm 0.004\%$  versus  $4.1 \pm 0.003\%$ ,  $P < 0.001$ ). The residual fluorescence at the end of a fourth trial (not shown) was taken as baseline. **d**, Differences in destaining kinetics with 10 Hz extracellular field stimulation ( $n = 120$  for FM1-84,  $n = 102$  for FM1-43,  $n = 89$  for FM2-10). Background fluorescence was determined following several applications of 90  $K^+$ /2  $Ca^{2+}$  solution. Inset shows the degree of FM1-43 destaining when FM2-10 fluorescence loss was  $\sim 60\%$  ( $47.5 \pm 1.3\%$  for 90  $K^+$  stimulation,  $52.3 \pm 1.1\%$  for 10 Hz,  $58.5 \pm 2.5\%$  for 1 Hz).



**Figure 4** A simple model accounts for the dependence of bouton destaining kinetics on dye species. **a**, Modelling of dye destaining during sustained stimulation. The observable fluorescence ( $F = S_0 + S_1 + S_2$ ) arises from dye in readily releasable ( $S_0$ ), fused ( $S_1$ ) and retrieved ( $S_2$ ) vesicles. **b**, Kinetic simulations of experimental data from Fig. 3b. In the best fit,  $\alpha = 0.5 s^{-1}$ ,  $\gamma = 0.014 s^{-1}$  and  $\beta = 0.167 s^{-1}$ , giving a time constant of fast endocytosis of 6 s;  $\delta$  was set by

measured rates of departitioning (Fig. 3a); 60% of the vesicles were assumed readily releasable. **c**, Simulation of a 'kink' by assuming six consecutive steps for vesicle recycling ( $\alpha = 0.5 s^{-1}$ ,  $\beta = 0.5 s^{-1}$ ,  $\delta = 0.4 s^{-1}$ , and each  $\gamma = 0.142 s^{-1}$ ). The dye content of the releasable pool ( $S_0$ ) was set at an initial value of 80%, and the intermediate states leading up to  $S_0$  were initially evenly populated.



**Figure 5** Effects of staurosporine and elevated  $[Ca^{2+}]_o$ . **a, b**, Ensemble averages of FM1-43 (**a**) and FM2-10 (**b**) destaining profiles from boutons pretreated for 1 h with 2  $\mu$ M staurosporine (top,  $n = 95$  for FM1-43,  $n = 98$  for FM2-10) or DMSO only (bottom). Smooth curves were simulated by changing only the endocytotic time constant ( $\beta^{-1}$ ), from 6 s (control) to 1.7 s (staurosporine). The broken line was simulated by slowing exocytosis threefold. **c**, Slowing of FM2-10 destaining kinetics by elevating  $[Ca^{2+}]_o$  from 1 to 8 mM. The model fitted the data when the endocytotic time constant was reduced from 8 s to 1.7 s, whereas a change of the exocytotic rate  $\alpha$  (from 0.5 s<sup>-1</sup> to 0.125 s<sup>-1</sup>) did not produce a good fit (broken trace). **d**, Summary of destaining experiments with elevated  $[Ca^{2+}]_o$ . Each data point indicates the mean fluorescence loss at 15 s (~3 times the longest departitioning time constant) following 90 K<sup>+</sup> stimulation.

modulatory control seems to be directed towards the kinetics of endocytosis.

The effects of elevated  $Ca^{2+}$  on endocytosis are not clear. In secretory cells, increases in internal  $Ca^{2+}$  are generally found to accelerate endocytosis<sup>9,10,24–26</sup>, although one study reports that high  $[Ca^{2+}]$  has the opposite effect<sup>27</sup>. Endocytosis at *Drosophila* neuromuscular junctions<sup>28</sup> shows an absolute requirement for  $Ca^{2+}$ , but no requirement for  $[Ca^{2+}]_o$  was found at hippocampal boutons<sup>4</sup>. Accordingly, we explored the effects of varying  $[Ca^{2+}]_o$  under conditions where fast endocytosis was prominent (Fig. 5c, d). Raising  $[Ca^{2+}]_o$  from 1 to 8 mM significantly decreased the initial release of FM2-10 (Fig. 5c, d), corresponding to increasing the rate of fast endocytosis from  $\tau \sim 8$  s to  $\tau \sim 2$  s. The effect is opposite to that expected for the  $Ca^{2+}$  dependence of exocytosis, which dominates at lower  $[Ca^{2+}]_o$ . The slower probes FM1-43 and FM1-84 were also increasingly retained as  $[Ca^{2+}]_o$  was elevated (Fig. 5d), although the degree of retention levelled off at ~60% (~40% fluorescence loss). Dye trapping would be expected to reach a ceiling once the rate of fast vesicle retrieval greatly exceeded the rate of dye departitioning. The remaining 40% of untrapped fluorescence suggests that a substantial fraction of vesicles may be shunted away from rapid endocytosis into a slow endocytotic pathway<sup>23</sup> where vesicles completely destain while awaiting retrieval (Fig. 1).

Our results demonstrate that rapid endocytosis of vesicles occurs at synapses between hippocampal neurons. The retrieval was swift enough to prevent complete loss of FM1-43 during an initial round of exocytosis, allowing subsequent dye release to appear as a delayed burst. The incompleteness of the initial dye loss suggests that free lateral diffusion of dye from vesicle to plasmalemma may be restricted by a barrier, as might arise from a fusion pore<sup>29</sup>. FM dyes with longer hydrocarbon tails were retained to increasing degrees in proportion to their measured departitioning rates. The series of dyes provided kinetic information presently unattainable with capacitance measurements, and distinct from that derived from synaptic currents, which mark precisely the moment of fusion

pore opening and transmitter efflux but give no direct information about the timing of endocytosis. Our estimates of the endocytotic time constant ranged from ~6 s under normal recording conditions to <2 s under modulation, an order of magnitude faster than previous estimates at hippocampal synapses<sup>4</sup>. Indeed, we may have underestimated the maximum speed of rapid endocytosis because the kinetic model lumps it together with any slow internalization working in parallel. The common occurrence of delayed bursts of destaining supports the idea that a large percentage of internalized vesicles recycle intact<sup>19</sup>, possibly by a 'kiss and run' mechanism<sup>16,17</sup> or by clathrin-mediated, non-endosomal retrieval<sup>20</sup>.

The rapid endocytosis at the higher stimulus frequencies (Fig. 3d) fits well with results from an earlier study<sup>4</sup> showing a sharp drop in the ability of hippocampal nerve terminals to take up FM1-43 just 5 s after the cessation of 10 Hz stimulation. The ability of staurosporine to inhibit FM dye release<sup>18,23</sup> can be explained by a decrease in the time the vesicular membrane is externally accessible<sup>18</sup> relative to that in the absence of drug<sup>30</sup> (Fig. 5a, b). The lack of effect of staurosporine on antibody uptake<sup>23</sup> also makes sense if this uptake largely involves a slow internalization pathway rather than fast retrieval.

Our experiments suggest that the rate of rapid endocytosis increases with stimulus strength<sup>10,12</sup> and elevated  $[Ca^{2+}]_o$ , bringing hippocampal nerve terminals in line with non-secretory systems<sup>10</sup>. Modulation of endocytic rate is clearly advantageous in achieving a balance between vesicular delivery and recycling, and may be important for the control of vesicle availability. The finding that nerve terminals of CNS neurons can make use of efficient mechanisms for rapid vesicular retrieval suggests new mechanisms for the way synaptic communication in the brain might be regulated. □

## Methods

**Cell culture.** Hippocampal neurons from cortical regions CA1–CA3 of Sprague-Dawley rats 2–3 days old were cultured according to previous protocols<sup>7,21</sup> with minor modifications. Cultures were used between 2 and 3 weeks *in vitro*.

**Dye loading and destaining.** A modified Tyrode solution (in mM: 150 NaCl, 4 KCl, 2 MgCl<sub>2</sub>, 2 CaCl<sub>2</sub>, 10 glucose, 10 HEPES, pH 7.4) was used as the extracellular medium for all experiments unless stated. The Tyrode solution always contained 10  $\mu$ M CNQX to prevent recurrent action potentials. Synaptic boutons were loaded with FM dyes (Molecular Probes) in the presence of 45 mM K<sup>+</sup> and 2 mM Ca<sup>2+</sup> for 90 s at room temperature. Individual boutons were imaged after 15 min perfusion with dye-free Tyrode's solution. Dye concentrations (2  $\mu$ M FM1-84, 4  $\mu$ M FM1-43 and 200  $\mu$ M FM2-10) were chosen to give similar signal intensities. Solution exchanges for staining and 90 K<sup>+</sup> stimulation were achieved by the fast movement (<200 ms) of an array of microcapillaries using an MPC-100 programmable micromanipulator (Sutter Instruments). Flow rate was 1 ml min<sup>-1</sup>. Extracellular field stimulation was achieved by 20 mA, 1 ms current pulses through two parallel platinum wires separated by 0.8–1.0 cm and positioned around the area of interest. Increasing stimulus frequency over the range 1–10 Hz sharply increased the rate of FM dye loss, with no hint of the fall in exocytotic activity found at 40 Hz at frog neuromuscular junctions<sup>30</sup>.

**Optical measurement and analysis.** A computer-controlled optical switch (DX-1000, Stanford Photonics) was coupled to the epifluorescence port of an inverted Nikon Diaphot TMD microscope by using a liquid light guide. For fluorescence imaging we used a Nikon  $\times 40$ , 1.3 N.A. oil-immersion objective lens. For each data point, FM dyes were excited at 480 to 40 nm (505 DCLP, 535/50 BP) for 66 ms, and two consecutive frames acquired with an intensified progressive line scan CCD camera (Stanford Photonics). With a study cycle of two frames per second, we measured a photo-bleaching time constant of ~1,000 s for all dyes. Accordingly, no corrections of the data for photobleaching were made. Images were digitized using board and software from Axon Instruments. Circular regions of interest ~1.5–2.0  $\mu$ m in diameter were defined around the brightness centre of mass of fluorescence spots. Pixel intensities within a region of interest were averaged for each set of two consecutive frames to give one data point. If a displacement of a bouton was observed in subsequent frames, the entire data set was discarded. The relative

loss of fluorescence for a given bouton was calculated as the difference between the average of 10 data points of baseline before a stimulus and the average of the last 10 data points of the time-lapse sequence.

**Modelling.** Exocytosis of releasable vesicles under a sustained stimulus was lumped into a single rate  $\alpha$  (see Fig. 4a). In the fused state, dye could either depart with rate  $\delta$  or be taken up again by means of fast endocytosis with rate  $\beta$ . Freshly endocytosed vesicles were recycled back to the readily releasable pool either with a single rate  $\gamma$  (Fig. 4b) or through a series of  $n$  consecutive steps, each having a rate constant  $\gamma/n$  (Fig. 4c). The variables  $S_x$  represent the relative amounts of dye incorporated into vesicle membrane pools, with subscripts 0, 1 and 2 denoting readily releasable, fused and retrieved vesicles. The state vector  $S = (S_0, S_1, S_2)^T$  represents the filling of the three states, and its evolution in time is described by the matrix equation

$$\frac{dS}{dt} = \begin{bmatrix} -\alpha & 0 & \gamma \\ \alpha & -(\beta + \delta) & 0 \\ 0 & \beta & -\gamma \end{bmatrix} S$$

with the sum of the states  $S_0$ ,  $S_1$  and  $S_2$  corresponding to the observable fluorescence. The amount of retrieved dye can be obtained by integrating the flux through the transition  $\beta$  (Fig. 4a, bottom).

Received 2 April; accepted 14 April 1998.

- Harris, K. M. & Sultan, P. Variation in the number, location and size of synaptic vesicles provides an anatomical basis for the nonuniform probability of release at hippocampal CA1 synapses. *Science* **264**, 1387–1395 (1995).
- Schikorski, T. & Stevens, C. F. Quantitative ultrastructural analysis of hippocampal excitatory synapses. *J. Neurosci.* **17**, 5858–5867 (1997).
- Betz, W. J., Mao, F. & Smith, C. B. Imaging exocytosis and endocytosis. *Curr. Opin. Neurobiol.* **6**, 365–371 (1996).
- Ryan, T. A., Smith, S. J. & Reuter, H. The timing of synaptic vesicle endocytosis. *Proc. Natl Acad. Sci. USA* **93**, 5567–5571 (1996).
- Cremona, O. & De Camilli, P. Synaptic vesicle endocytosis. *Curr. Opin. Neurobiol.* **7**, 323–330 (1997).
- Ryan, T. A. *et al.* The kinetics of synaptic vesicle recycling measured at single presynaptic boutons. *Neuron* **11**, 713–724 (1993).
- Liu, G. & Tsien, R. W. Properties of synaptic transmission at single hippocampal synaptic boutons. *Nature* **268**, 1624–1628 (1995).
- Wu, L. G. & Betz, W. J. Nerve activity but not intracellular calcium determines the time course of endocytosis at the frog neuromuscular junction. *Neuron* **17**, 769–779 (1996).
- Thomas, P., Lee, A. K., Wong, J. G. & Almers, W. A triggered mechanism retrieves membrane in seconds after  $Ca^{2+}$ -stimulated exocytosis in single pituitary cells. *J. Cell Biol.* **124**, 667–675 (1994).
- Smith, C. & Heuser, E. Multiple forms of endocytosis in bovine adrenal chromaffin cells. *J. Cell Biol.* **139**, 885–894 (1997).
- Engisch, K. L. & Nowycky, M. Compensatory and excess retrieval: two types of endocytosis following single step depolarizations in bovine adrenal chromaffin cells. *J. Physiol. (Lond.)* **506**, 591–608 (1998).
- Henkel, A. W. & Almers, W. Fast steps in exocytosis and endocytosis studied by capacitance measurements in endocrine cells. *Curr. Opin. Neurobiol.* **6**, 350–357 (1996).
- Heuser, J. E. & Reese, T. S. Evidence for recycling of synaptic vesicle membrane during transmitter release at the frog neuromuscular junction. *J. Cell Biol.* **57**, 315–344 (1973).
- Miller, T. M. & Heuser, J. E. Endocytosis of synaptic vesicle membrane at the frog neuromuscular junction. *J. Cell Biol.* **98**, 685–698 (1984).
- Koenig, J. H. & Ikeda, K. Synaptic vesicles have two distinct recycling pathways. *J. Cell Biol.* **135**, 797–808 (1996).
- Ceccarelli, B. & Hurlbut, W. P.  $Ca^{2+}$ -dependent recycling of synaptic vesicles at the frog neuromuscular junction. *J. Cell Biol.* **87**, 297–303 (1980).
- Fesce, G., Grohovaz, E., Valtorta, F. & Meldolesi, J. Neurotransmitter release: fusion or “kiss and run”? *Trends Cell Biol.* **4**, 1–4 (1994).
- Henkel, A. W. & Betz, W. J. Staurosporine blocks evoked release of FM1-43 but not acetylcholine from frog motor nerve terminals. *J. Neurosci.* **15**, 8246–8258 (1995).
- Murthy, V. N. & Stevens, C. F. Synaptic vesicles retain their identity through the endocytotic cycle. *Nature* **392**, 497–501 (1998).
- Takei, K., Mundigl, O., Daniell, L. & De Camilli, P. The synaptic vesicle cycle: a single vesicle budding step involving clathrin and dynamin. *J. Cell Biol.* **133**, 1237–1250 (1996).
- Malgarelli, A. & Tsien, R. W. Glutamate-induced long-term potentiation of the frequency of miniature synaptic currents in cultured hippocampal neurons. *Nature* **357**, 134–139 (1992).
- Lagnado, L., Gomis, A. & Job, C. Continuous vesicle cycling in the synaptic terminal of retinal bipolar cells. *Neuron* **17**, 957–967 (1996).
- Kraszewski, K., Daniell, L., Mundigl, O. & De Camilli, P. Mobility of synaptic vesicles in nerve endings monitored by recovery from photobleaching of synaptic vesicle-associated fluorescence. *J. Neurosci.* **16**, 5905–5913 (1996).
- Parsons, T. D., Lenzi, D., Almers, W. & Roberts, W. M. Calcium-triggered exocytosis and endocytosis in an isolated presynaptic cell: capacitance measurements in saccular hair cells. *Neuron* **13**, 875–883 (1994).
- Burgoyne, R. D. Fast exocytosis and endocytosis triggered by depolarisation in single adrenal chromaffin cells before rapid  $Ca^{2+}$  current run-down. *Pflügers Arch.* **430**, 213–219 (1995).
- Plattner, H., Braun, C. & Hentschel, J. Facilitation of membrane fusion during exocytosis and exocytosis-coupled endocytosis and acceleration of “ghost” detachment in *Paramecium* by extracellular calcium. A quenched-flow/freeze-fracture analysis. *J. Membr. Biol.* **158**, 197–208 (1997).
- von Gersdorff, H. & Matthews, G. Inhibition of endocytosis by elevated internal calcium in a synaptic terminal. *Nature* **370**, 652–666 (1994).
- Ramaswami, M., Krishnan, K. S. & Kelly, R. B. Intermediates in synaptic vesicle recycling revealed by optical imaging of *Drosophila* neuromuscular junctions. *Neuron* **13**, 363–375 (1994).
- Zimmerberg, J., Blumenthal, R., Sarkar, D. P., Curran, M. & Morris, S. J. Restricted movement of lipid and aqueous dyes through pores formed by influenza hemagglutinin during cell fusion. *J. Cell Biol.* **127**, 1885–1894 (1994).
- Betz, W. J. & Bewick, G. S. Optical monitoring of transmitter release and synaptic vesicle recycling at the frog neuromuscular junction. *J. Physiol. (Lond.)* **460**, 287–309 (1993).

**Acknowledgements.** We thank S. J. Smith and T. L. Schwarz for critically reading the manuscript, members of the Tsien laboratory for comments, and J. Bergsman for discussions and suggestions. This work was supported by grants from NIMH, the Mathers Charitable Foundation, the McKnight Endowment Fund for Neuroscience (R.W.T.), and fellowships from the Boehringer Ingelheim Fonds (J.K.) and the American Heart Association California Affiliate (E.T.K.).

Correspondence and requests for materials should be addressed to R.W.T. (e-mail: rwtstien@leland.stanford.edu).

## Nitric oxide functions as a signal in plant disease resistance

Massimo Delledonne<sup>\*†‡</sup>, Yiji Xia<sup>\*‡§</sup>, Richard A. Dixon<sup>§</sup> & Chris Lamb<sup>\*</sup>

<sup>\*</sup> Plant Biology Laboratory, The Salk Institute, La Jolla, California 92037, USA

<sup>§</sup> Plant Biology Division, The Noble Foundation, Ardmore, Oklahoma 73402, USA

<sup>‡</sup> These authors contributed equally to this work.

Recognition of an avirulent pathogen triggers the rapid production of the reactive oxygen intermediates superoxide ( $O_2^-$ ) and hydrogen peroxide ( $H_2O_2$ )<sup>1</sup>. This oxidative burst drives cross-linking of the cell wall<sup>2</sup>, induces several plant genes involved in cellular protection and defence<sup>3,4</sup>, and is necessary for the initiation of host cell death in the hypersensitive disease-resistance response<sup>1,3</sup>. However, this burst is not enough to support a strong disease-resistance response<sup>4,5</sup>. Here we show that nitric oxide, which acts as a signal in the immune, nervous and vascular systems<sup>6</sup>, potentiates the induction of hypersensitive cell death in soybean cells by reactive oxygen intermediates and functions independently of such intermediates to induce genes for the synthesis of protective natural products. Moreover, inhibitors of nitric oxide synthesis compromise the hypersensitive disease-resistance response of *Arabidopsis* leaves to *Pseudomonas syringae*, promoting disease and bacterial growth. We conclude that nitric oxide plays a key role in disease resistance in plants.

The oxidative burst required for hypersensitive cell death in soybean cell suspensions inoculated with avirulent *P. syringae* pv. *glycinea* gives a sustained accumulation of 6–10  $\mu M$   $H_2O_2$  (ref. 7). Addition of the  $H_2O_2$ -generating system glucose (0.5 mM)/glucose oxidase (0.5 U ml<sup>-1</sup>) to these cell suspensions resulted in a steady-state accumulation of ~6  $\mu M$   $H_2O_2$  for >3 h, and hence closely mimics the oxidative burst induced by the avirulent pathogen. However, *in situ* generation of  $H_2O_2$  in this fashion, or equivalent generation of  $O_2^-$  by addition of xanthine/xanthine oxidase, induced only a relatively weak cell-death response (Fig. 1a).

In the immune system, reactive oxygen intermediates (ROI) often function together with nitric oxide (NO), for example in macrophage killing of bacteria and tumour cells<sup>6,8</sup>. We therefore tested whether NO might be the second factor required for a strong hypersensitive disease-resistance response (HR). Treatment of cell suspensions with 0.5 mM sodium nitroprusside generated a steady-state NO concentration of ~2  $\mu M$  (see below), which in the absence of ROI did not induce cell death. However, this NO donor markedly potentiated the induction of cell death by exogenous  $H_2O_2$  or  $O_2^-$  (Fig. 1a). The interaction was synergistic, with NO promoting a 5–10-fold increase in ROI-induced cell death. Potassium ferrocyanide, an analogue of sodium nitroprusside that does not release NO, had no effect on ROI-induced cell death.

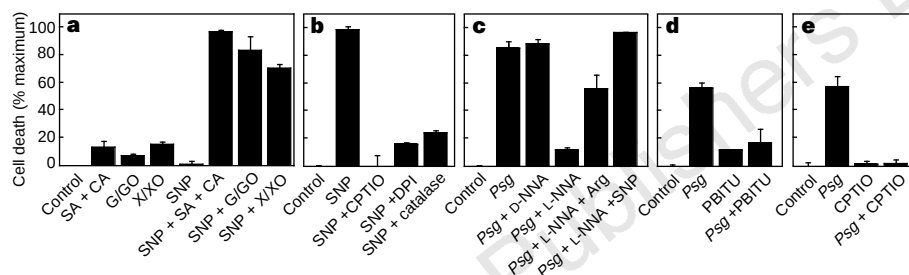
NO also potentiated cell death induced by endogenous ROI. Activation of the oxidative burst following pathogen recognition involves a protein kinase cascade<sup>3,9</sup>. Cantharidin, an inhibitor of

<sup>†</sup> Permanent address: Istituto di Genetica, Università Cattolica S.C., Piacenza, 29100 Italy.

type-2a protein phosphatases, partially activates this signalling pathway in the absence of pathogen avirulence factors<sup>3</sup>. Salicylic acid, which functions in signal amplification<sup>7</sup>, enhances this effect, resulting in the accumulation of  $>30 \mu\text{M}$   $\text{H}_2\text{O}_2$  (Fig. 2i; ref. 7). This massive oxidative burst caused only a weak induction of cell death, but simultaneous addition of nitroprusside strongly potentiated the response (Fig. 1a). Likewise, a yeast cell-wall elicitor, which rapidly stimulates a strong oxidative burst<sup>10</sup> (Fig. 2j), did not induce cell death without exogenous NO (data not shown). Rapid agitation of plant cells also causes ROI production<sup>11</sup> and, under these conditions, which gave a steady-state  $\text{H}_2\text{O}_2$  concentration of  $\sim 1 \mu\text{M}$  in the soybean cultures, nitroprusside induced cell death without added ROI (Fig. 1b). This response was blocked not only by the NO scavenger 2-4-carboxyphenyl-4,4,5,5-tetramethylimidazoline-

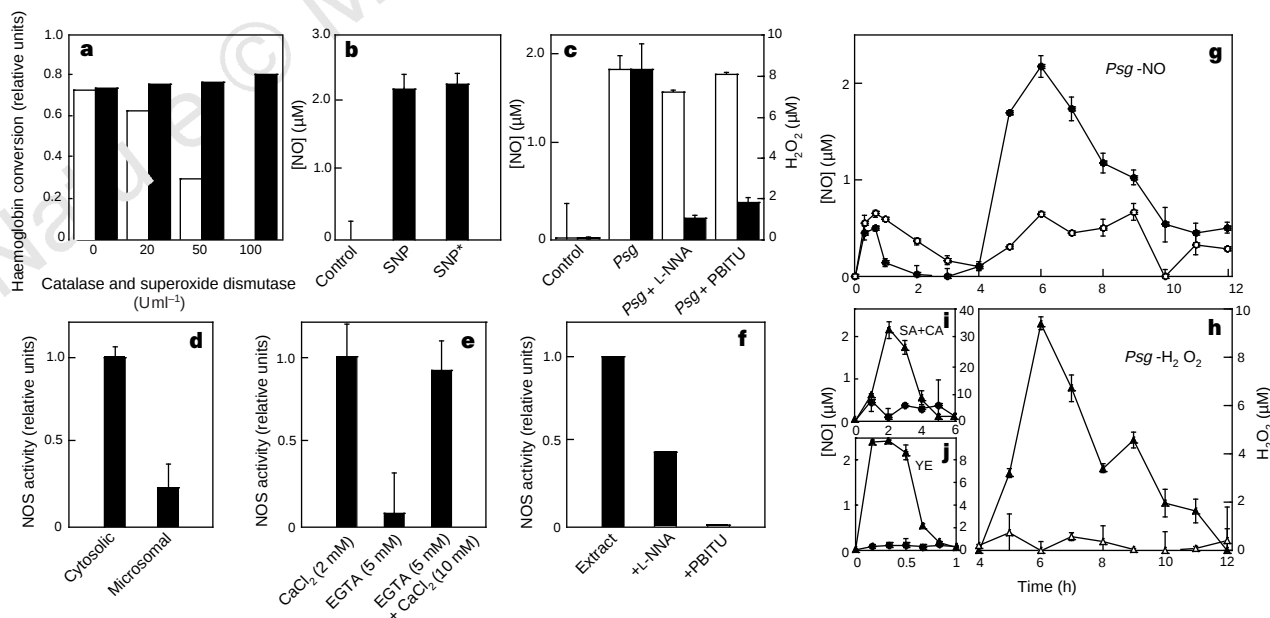
1-oxyl-3-oxide (CPTIO), but also by diphenylene iodonium, an inhibitor of the neutrophil NADPH oxidase that inhibits the plant oxidative burst<sup>3</sup>, and by catalase, which destroys  $\text{H}_2\text{O}_2$ , thereby confirming the involvement of endogenous ROI together with the added NO (Fig. 1b). Two other NO donors, *S*-nitroso-*N*-acetylpenicillamine and diethylamino NO, gave a similar response (data not shown).

The synergistic interaction between exogenous NO and ROI suggested that NO might have a physiological function in the HR. Endogenous NO was assayed by the conversion of haemoglobin to methaemoglobin<sup>12</sup>. ROI can interfere with this assay, and although haemoglobin conversion was observed with cells treated with yeast elicitor, pretreatment of the assay sample with superoxide dismutase plus catalase abolished this spurious signal (Fig. 2a). Neither yeast



**Figure 1** Function of NO in the induction of hypersensitive cell death in soybean cell suspension cultures. **a, b**, NO potentiation of ROI-induced cell death. **c, d**, NO function in *P. syringae* pv. *glycinea* (*Psg*)-induced hypersensitive cell death. In **b**, cell suspensions were rapidly agitated (100 r.p.m.) to stimulate ROI generation sufficient to maintain a steady-state  $\text{H}_2\text{O}_2$  concentration of  $\sim 1 \mu\text{M}$ . In **a, c-e**, cell suspensions were shaken slowly (60 r.p.m.) and the basal level of  $\text{H}_2\text{O}_2$  was  $<0.05 \mu\text{M}$ . Reagents were added as indicated at the following final concentrations:

50  $\mu\text{M}$  salicylic acid (SA), 4  $\mu\text{M}$  cantharidin (CA), 0.5 mM glucose + 0.5 U  $\text{ml}^{-1}$  glucose oxidase (G/GO), 0.5 mM xanthine + 0.5 U  $\text{ml}^{-1}$  xanthine oxidase (X/XO), 0.5 mM nitroprusside (SNP), 5 mM (**b**) and 50  $\mu\text{M}$  (**d**) CPTIO, 10  $\mu\text{M}$  diphenylene iodonium (DPI), 300 U  $\text{ml}^{-1}$  catalase, 200  $\mu\text{M}$  L-NNA or D-NNA, 1 mM PBITU, 2 mM arginine. Cell death was assayed after 24 h by spectrophotometric determination of Evans blue uptake relative to that in equivalent cell suspensions killed by ethanol.



**Figure 2** NO production in soybean cell suspensions. **a**, Validation of NO assay. The effects of pretreatment of assay samples with the indicated activities of superoxide dismutase and catalase on haemoglobin conversion were measured in samples from cells induced with yeast elicitor for 20 min (white bars) or *P. syringae* pv. *glycinea* (*Psg*) for 6 h (black bars). **b**, NO accumulation in cells 2 h after treatment with 0.5 mM nitroprusside (SNP). Asterisk denotes cells that were rapidly shaken to give mechanically induced accumulation of ROI. **c**, Effect of 0.2 mM L-NNA or 1 mM PBITU on *Psg* induction of  $\text{H}_2\text{O}_2$  (white bars) and NO (black

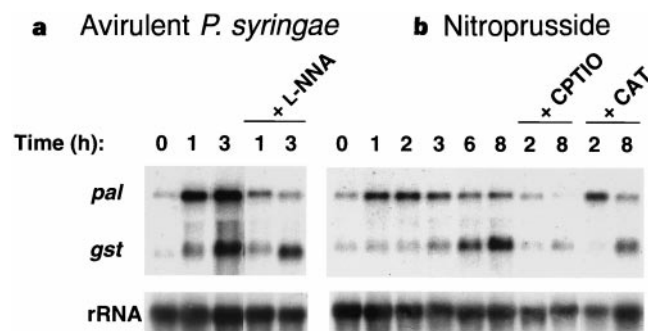
bars) accumulation after 6 h. **d**, Distribution of NOS activity between cytosolic and microsomal fractions of soybean whole-cell extracts. Relative activities are presented on a cell-equivalent basis. **e**, Calcium dependence of cytosolic NOS activity. **f**, Effect of 0.2 mM L-NNA or 1 mM PBITU on NOS activity in whole-cell extracts. **g-j**, NO (circles) and  $\text{H}_2\text{O}_2$  (triangles) accumulation in soybean cell suspensions treated with: **g, h**, *Psg* (filled symbols: avirulent strain; open symbols: isogenic virulent strain); **i**, cantharidin (CA, 1  $\mu\text{M}$ ) + salicylic acid (SA, 50  $\mu\text{M}$ ); **j**, yeast elicitor (YE, 100  $\mu\text{g}$  glucose equivalents  $\text{ml}^{-1}$ ).



elicitor nor salicylic acid plus cantharidin stimulated NO accumulation, as measured by haemoglobin assay after superoxide dismutase and catalase pretreatment (Fig. 2i, j), consistent with the inability of these stimuli to trigger cell death despite activation of a strong oxidative burst. In contrast, inoculation of soybean cell suspensions with *P. syringae* pv. *glycinea* stimulated NO production, as measured by assay of samples depleted for ROI by pretreatment with superoxide dismutase plus catalase (Fig. 2a, g). A rapid, relatively weak stimulation by both avirulent and virulent strains was followed by several-fold greater NO production specifically in cells inoculated with the avirulent strain, concomitant with the avirulence gene-dependent oxidative burst (Fig. 2g, h). The maximal accumulation of 2  $\mu$ M NO and of 10  $\mu$ M H<sub>2</sub>O<sub>2</sub> closely matched their respective levels in the experiments with nitroprusside and glucose/glucose oxidase (Fig. 2b), and hence would be sufficient to account for the induction of hypersensitive cell death by avirulent *P. syringae*.

In animals, NO is generated by the reaction arginine → citrulline + NO, which is catalysed by NO synthase (NOS)<sup>6</sup>. Plants also have NOS activity<sup>13,14</sup>, and in our soybean cell extracts NOS activity was calcium-dependent and found primarily in the cytosolic fraction (Fig. 2d, e). The NOS inhibitors N<sup>ω</sup>-nitro-L-arginine (L-NNA) and S,S'-1,3-phenylene-bis(1,2-ethanediyl)-bis-isothiourea (PBITU)<sup>15,16</sup>, which inhibits this reaction in soybean cell extracts (Fig. 2f), markedly inhibited (at concentrations comparable to those used with animal cells) the accumulation of endogenous NO in bacterially induced cells, but had no effect on the accumulation of H<sub>2</sub>O<sub>2</sub> (Fig. 2c). Both PBITU and L-NNA, but not the inactive enantiomer D-NNA, blocked bacterial induction of hypersensitive cell death (Fig. 1c, d). L-NNA inhibition could be partially competed with exogenous arginine or completely bypassed with nitroprusside (Fig. 1c). Moreover, the induction of hypersensitive cell death by avirulent *P. syringae* was also inhibited by the NO scavenger CPTIO (Fig. 1e), which blocks NO accumulation by a different mechanism.

H<sub>2</sub>O<sub>2</sub> drives oxidative crosslinking of tyrosine-rich cell-wall structural proteins<sup>2</sup> and the oxidative burst also induces protective cellular genes encoding glutathione S-transferase (*gst*) and glutathione peroxidase<sup>3</sup>. In soybean cells, however, ROI are not the primary signals for rapid induction of defence genes encoding enzymes of phenylpropanoid biosynthesis involved in the synthesis of antibiotics, lignin and salicylic acid<sup>15</sup>. Whereas L-NNA had little effect on the initial induction of *gst* by avirulent *P. syringae*, the accumulation of transcripts encoding phenylalanine ammonia-lyase (*pal*), the first enzyme of the phenylpropanoid pathway<sup>17</sup>, was markedly reduced (Fig. 3). Bacterial induction of transcripts

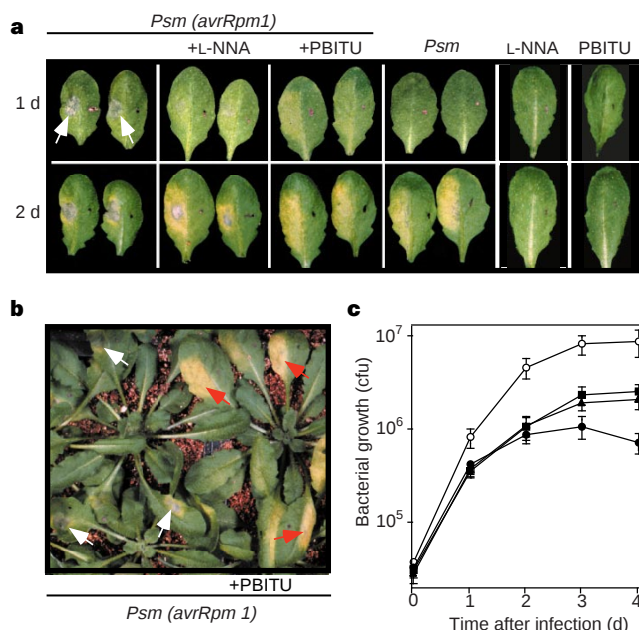


**Figure 3** NO induction of *pal*. Northern blot analysis of **a**, the effects of the NO synthase inhibitor L-NNA on *pal* and *gst* induction by avirulent *P. syringae* pv. *glycinea*, and **b**, NO induction of *pal* and *gst*. RNA loading was checked by hybridization of rRNA. Reagents were added as indicated at the following final concentrations: 0.5 mM nitroprusside, 200  $\mu$ M L-NNA, 50  $\mu$ M CPTIO, 300 U ml<sup>-1</sup> catalase (CAT).

encoding chalcone synthase (*chs*), the first enzyme of the branch specific for flavonoids and isoflavonoid-derived antibiotics<sup>17</sup>, showed a similar dependence on endogenous NO (data not shown). Moreover, sodium nitroprusside caused a rapid induction of *pal* and *chs* transcripts that was blocked by CPTIO but not by catalase. In contrast, nitroprusside caused only a slow accumulation of *gst* transcripts and this was sensitive to catalase as well as to CPTIO (Fig. 3).

To test whether NO functions *in planta*, we examined the effects of inhibitors of NO accumulation in a genetically well defined host-pathogen interaction. Infiltration of leaves of *Arabidopsis* ecotype Col-0, which contains the *RPM1* resistance gene, with *P. syringae* pv. *maculicola* carrying the corresponding *avrRpm1* avirulence gene, results in a localized HR lesion which is first apparent ~1 d after inoculation<sup>18</sup>. Co-infiltration of the NOS inhibitors L-NNA or PBITU blocked this early HR and promoted the spreading chlorosis observed in the compatible interaction with the isogenic virulent bacteria lacking *avrRpm1* (Fig. 4a, b). L-NNA and PBITU alone did not cause disease symptoms (Fig. 4a) and had no effect on axenic bacterial growth (data not shown). Development of disease symptoms in leaves blocked for NO accumulation was accompanied by enhanced bacterial growth, although not to the extent seen with the isogenic virulent strain (Fig. 4c).

In the vertebrate native immune system, NO potentiates the antimicrobial and cytotoxic activities of ROI<sup>6,8</sup>. We have shown by several independent approaches that NO also functions in the plant HR and appears to be critical for preventing pathogen spread from the inoculation site. Moreover, our results reveal a striking complementarity between NO and ROI signal functions, involving not only synergistic induction of hypersensitive cell death, but also independent activation of complementary gene sets. This arrangement provides for flexible deployment of specific defences and binary control of progression to hypersensitive cell death.



**Figure 4** Function of NO in the hypersensitive disease-resistance response: effects of L-NNA and PBITU on: **a**, **b**, visible symptoms, and **c**, bacterial growth. *Psm* represents virulent *P. syringae* pv. *maculicola* (open circles); *Psm* (*avrRpm1*), the isogenic avirulent strain of *P. syringae* pv. *maculicola* inoculated alone (filled circles), with L-NNA (triangles) or with PBITU (squares). PBITU was 1 mM in all experiments, L-NNA was 0.1 mM in **a**, 0.2 mM in **c**. Red arrows mark disease; white arrows, HR. In **c**, data points are the mean and standard deviation of three replicates, each from three leaf discs. These experiments were repeated three times with similar results.

In animals, many physiological effects of NO are mediated by cyclic GMP, following direct activation of guanylyl cyclase<sup>6,8</sup>. In plants, cGMP functions in the phytochrome-mediated light induction of *chs* associated with the formation of phenylpropanoid-derived pigments<sup>19</sup>. Recent results indicate that cGMP also functions in the induction of *pal* by exogenous NO (ref. 29), associated with the synthesis of phenylpropanoid products with defence functions<sup>20</sup>. Moreover, ROI induce hypersensitive cell death by raising cytosolic calcium<sup>21</sup> and NO can reduce the threshold for calcium signalling by modulation of cGMP-gated ion channels<sup>22</sup>. Hence, NO may activate cGMP pathways for both defence gene induction and potentiation of ROI-induced cell death.

An enzyme system similar to the neutrophil NADPH oxidase apparently contributes to the oxidative burst in plants<sup>23,24</sup>. Although NOS homologues have not yet been reported in plants, a plant complementary DNA sequence with pronounced similarity to PIN, a protein inhibitor of NOS, has been identified<sup>25</sup>, suggesting that functional parallels between the native immune system and the plant HR may be maintained by conserved molecular mechanisms. □

## Methods

Soybean (*Glycine max* cv. Williams 82) cell cultures, virulent *P. syringae* pv. *glyciniae* race 4 and the isogenic strain carrying the *avrA* avirulence gene, which interacts with the soybean resistance gene *Rgm2* (ref. 26) were grown as described<sup>13,7</sup>. Three days after subculture, aliquots of the soybean cell suspension were transferred to 12-well tissue-culture plates (1 ml per well) agitated at 60 or 100 r.p.m. Bacteria ( $10^8$  colony forming units (CFU) per ml) and reagents at the indicated concentrations were added simultaneously as appropriate. The growth of *A. thaliana* ecotype Col-0, infiltration of leaves with virulent *P. syringae* pv. *maculicola* or the isogenic avirulent strain carrying *avrRpm1* ( $5-10 \mu\text{l}$  of a bacterial suspension containing  $5 \times 10^6$  CFU  $\text{ml}^{-1}$ ), together with PBITU or L-NNA as indicated, and measurement of bacterial growth *in planta* have all been described<sup>27</sup>.

Cell death was assayed by spectrophotometric measurement of the uptake of Evans blue stain<sup>7</sup>, and  $\text{H}_2\text{O}_2$  accumulation by fluorescence determination of scopoletin destruction<sup>3</sup>. For NO assay, soybean cell suspensions were incubated with 100 U catalase and 100 U superoxide dismutase for 5 min to remove ROI before addition of oxyhaemoglobin ( $5 \mu\text{M}$  final). After 2 min incubation, NO was quantified by spectrophotometric measurement of the conversion of oxyhaemoglobin to methaemoglobin<sup>12</sup>. Each cell death,  $\text{H}_2\text{O}_2$  and NO datum point is the mean and standard deviation of three to five replicates. NOS activity was assayed by the conversion of L-[ $^3\text{H}$ ]-arginine to L-[ $^3\text{H}$ ]-citrulline<sup>28</sup> in whole-cell extracts, and microsomal and cytosolic fractions<sup>24</sup>. Northern blot hybridization was done as described<sup>3</sup>. Cantharidin, diphenylene iodonium, glucose oxidase and PBITU were from Calbiochem; CPTIO, diethylamino NO, S-nitroso-N-acetylpenicillamine were from Molecular Probes; all other reagents and enzymes were from Sigma.

Received 5 December 1997; accepted 1 June 1998.

1. Lamb, C. & Dixon, R. A. The oxidative burst in plant disease resistance. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 251–275 (1997).
2. Bradley, D. J., Kjellbom, P. & Lamb, C. J. Elicitor- and wound-induced oxidative cross-linking of a proline-rich plant cell wall protein: A novel, rapid defense response. *Cell* **70**, 21–30 (1992).
3. Levine, A., Tenhaken, R., Dixon, R. A. & Lamb, C.  $\text{H}_2\text{O}_2$  from the oxidative burst orchestrates the plant hypersensitive response. *Cell* **79**, 583–593 (1994).
4. Jabs, T., Tschöpe, M., Colling, C., Hahlbrock, K. & Scheel, D. Elicitor-stimulated ion fluxes and  $\text{O}_2^-$  from the oxidative burst are essential components in triggering defense gene activation and phytoalexin synthesis in parsley. *Proc. Natl Acad. Sci. USA* **94**, 4800–4805 (1997).
5. Glazener, J. A., Orlandi, E. W. & Baker, J. C. The active oxygen response of cell suspensions is not sufficient to cause hypersensitive cell death. *Plant Physiol.* **110**, 759–763 (1996).
6. Schmidt, H. H. W. & Walter, U. NO at work. *Cell* **78**, 919–925 (1994).
7. Shirasu, K., Nakajima, H., Rajasekhar, V. K., Dixon, R. A. & Lamb, C. Salicylic acid potentiates an agonist-dependent gain control that amplifies pathogen signals in the activation of defense mechanisms. *Plant Cell* **9**, 261–270 (1997).
8. Nathan, C. Natural resistance and nitric oxide. *Cell* **82**, 873–876 (1995).
9. Chandra, S., Martin, G. B. & Low, P. S. The Pto kinase mediates a signaling pathway leading to the oxidative burst in tomato. *Proc. Natl Acad. Sci. USA* **93**, 13393–13397 (1996).
10. Guo, Z.-J. & Ohta, Y. Effect of ethylene biosynthesis on the accumulation of 6-methoxymellein induced by elicitors in carrot cells. *J. Plant Physiol.* **144**, 700–704 (1994).
11. Yahraus, T., Chandra, S., Legendre, L. & Low, P. S. Evidence for a mechanically induced oxidative burst. *Plant Physiol.* **109**, 1259–1266 (1995).
12. Murphy, M. E. & Noack, E. Nitric oxide assay using hemoglobin method. *Meth. Enzymol.* **233**, 241–250 (1994).

13. Cueto, M. *et al.* Presence of nitric oxide synthase activity in roots and nodules of *Lupinus albus*. *FEBS Lett* **398**, 159–164 (1996).
14. Ninnemann, H. & Maier, J. Indications for the occurrence of nitric oxide synthases in fungi and plants and the involvement in photooxidation of *Neurospora crassa*. *Photochem. Photobiol.* **64**, 393–398 (1996).
15. Griffith, O. W. & Stuehr, D. J. Nitric oxide synthases: Properties and catalytic mechanism. *Annu. Rev. Physiol.* **57**, 707–736 (1995).
16. Garvey, E. P. *et al.* Potent and selective inhibition of human nitric oxide synthases. *J. Biol. Chem.* **269**, 26669–26676 (1994).
17. Dixon, R. A. & Paiva, N. Stress-induced phenylpropanoid metabolism. *Plant Cell* **7**, 1085–1097 (1995).
18. Debener, T., Lehnackers, H., Arnold, M. & Dangel, J. L. Identification and molecular mapping of a single *Arabidopsis thaliana* locus determining resistance to a phytopathogenic *Pseudomonas syringae* isolate. *Plant J.* **1**, 289–302 (1991).
19. Bowler, C., Neuhaus, G., Yamagata, H. & Chua, N.-H. Cyclic GMP and calcium mediate phytochrome phototransduction. *Cell* **77**, 73–81 (1994).
20. Noritake, T., Kawakita, K. & Doke, N. Nitric oxide induces phytoalexin accumulation in potato tuber tissues. *Plant Cell Physiol.* **37**, 113–116 (1996).
21. Levine, A., Pennell, R. I., Alvarez, M. E., Palmer, R. & Lamb, C. Calcium-stimulated apoptosis in a plant hypersensitive disease resistance response. *Curr. Biol.* **6**, 427–437 (1996).
22. Berridge, M. J. A tale of two messengers. *Nature* **365**, 388–389 (1993).
23. Groom, Q. J. *et al.* *rbhoA*, a rice homologue of the mammalian gp91<sup>phox</sup> respiratory burst oxidase gene. *Plant J.* **10**, 515–522 (1996).
24. Keller, T. *et al.* A plant homolog of the neutrophil NADPH oxidase gp91<sup>phox</sup> subunit gene encodes a plasma membrane protein with  $\text{Ca}^{2+}$ -binding domains. *Plant Cell* **10**, 255–266 (1998).
25. Jaffrey, S. R. & Snyder, S. H. PIN: an associated protein inhibitor of neuronal nitric oxide synthase. *Science* **274**, 774–777 (1996).
26. Keen, N. T. & Buzzell, R. I. New disease resistance genes in soybean against *Pseudomonas syringae* pv. *glyciniae*: Evidence that one of them interacts with a bacterial elicitor. *Theor. Appl. Genet.* **81**, 133–138 (1991).
27. Cameron, R. K., Dixon, R. A. & Lamb, C. J. Biologically induced systemic acquired resistance in *Arabidopsis thaliana*. *Plant J.* **5**, 715–725 (1994).
28. Hevel, J. M. & Marletta, M. A. Nitric oxide synthase assays. *Meth. Enzymol.* **23**, 251–258 (1994).
29. Durner, J., Wendehenne, D. & Klessig, D. F. Defense gene induction in tobacco by nitric oxide, cyclic GMP and cyclic ADP ribose. *Proc. Natl Acad. Sci. USA* (in the press).

**Acknowledgements.** Y.X. is a Noble/Salk postdoctoral fellow. This research was supported by grants to M.D. and C.L. from the Italian National Research Council and the Noble Foundation, respectively.

Correspondence and requests for materials should be addressed to C.L. (e-mail: lamb@salk.edu).

## Transformation of primary human endothelial cells by Kaposi's sarcoma-associated herpesvirus

Ornella Flore\*†, Shahin Rafii‡, Scott Ely\*, John J. O'Leary\*§, Elizabeth M. Hyjek\* & Ethel Cesarman\*

\* Department of Pathology, ‡ Division of Hematology–Oncology, Department of Medicine, Cornell University Medical College, New York, New York 10021, USA

† Dipartimento di Medicina Interna, Università di Cagliari, Cagliari, 09100, Italy

§ Departments of Pathology, The Coombe Women's Hospital and St James's Hospital Dublin, and Trinity College Dublin, Ireland

Kaposi's sarcoma-associated herpesvirus (KSHV), or human herpesvirus 8, is invariably present in Kaposi's sarcoma lesions<sup>1,2</sup>. KSHV contains several viral oncogenes and serological evidence suggests that KSHV infection is necessary for the development of Kaposi's sarcoma, but cellular transformation by this virus has not so far been demonstrated. KSHV is found in the microvascular endothelial cells in Kaposi's sarcoma lesions and in the spindle 'tumour' cells<sup>3,4</sup>, which are also thought to be of endothelial origin. Here we investigate the biological consequences of infecting human primary endothelial cells with purified KSHV particles. We find that infection causes long-term proliferation and survival of these cells, which are associated with the acquisition of telomerase activity and anchorage-independent growth. KSHV was present in only a subset of cells, and paracrine mechanisms were found to be responsible for the survival of uninfected cells. Their survival may have been mediated by upregulation of a receptor for vascular endothelial growth factor. Our results indicate that transformation of endothelial cells by KSHV, as

well as paracrine mechanisms that are induced by this virus, may be critical in the pathogenesis of Kaposi's sarcoma.

To investigate whether KSHV can infect primary human endothelial cells (ECs), we purified and concentrated viral particles from a KSHV-positive/Epstein-Barr (EBV)-negative cell line (BC-3)<sup>5</sup> derived from a primary effusion lymphoma. BC-3 cells were induced with the phorbol ester tetradecanoyl phorbol acetate (TPA) to activate the viral lytic cycle and the release of viral particles into the medium<sup>6,7</sup>. Early-passage bone-marrow microvascular ECs (BMECs)<sup>8</sup> were infected in duplicate with purified KSHV particles (5–10 genome equivalents per cell) and cultured in the presence of vascular endothelial growth factor A (VEGF). Two days post-infection, the cells displayed a clear cytopathic effect, whereas the uninfected monolayers remained unchanged. The surviving cells continued to proliferate, and after 15 days the monolayers were replaced by cells with a spindle-shape morphology, which were positive for KSHV DNA sequences (Fig. 1).

Two months after infection, KSHV-infected cells were still proliferating and still contained the KSHV genome. ECs in culture depend on VEGF for survival, and without it they stop proliferating and undergo apoptosis<sup>9</sup>. To determine whether infected BMECs required VEGF for growth, the VEGF concentration in the medium was reduced to sub-physiological concentrations (1–2 ng ml<sup>-1</sup>). Rather than proliferating at low VEGF concentrations, KSHV underwent lytic replication. Within 36 hours, multinucleated cells and foci of cell lysis showing a cytopathic effect typical of herpesvirus were visible throughout the culture (Fig. 1b). To confirm that lytic viral replication had occurred, supernatants from infected BMECs were treated in the same way as those from BC-3 cells, and the purified virus was used to infect primary human umbilical vein ECs (HUVECs) in triplicate. These serially KSHV-infected cells had the same features as infected BMECs. After prolonged growth in culture, the surviving cells also showed increased proliferation compared with uninfected controls, and were positive for KSHV DNA sequences by, as tested by polymerase chain reaction (PCR)

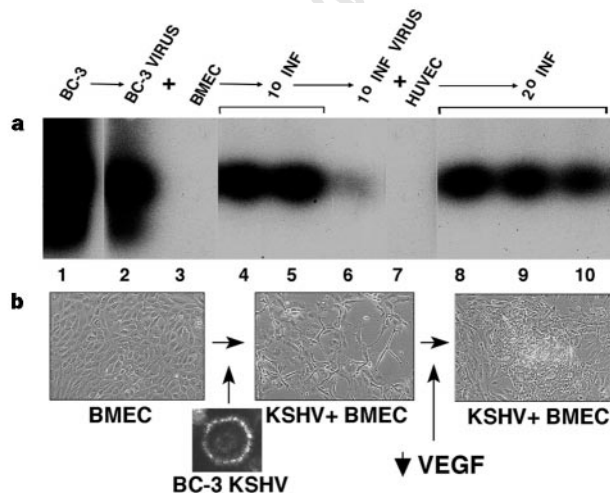
(Fig. 1a). The infected BMECs and HUVECs are still growing in culture after 14 and 12 months, respectively, whereas uninfected controls suffered senescence and died about three weeks after mock infection. BC-3 cells, viral preparations and infected cultures are negative for EBV, cytomegalovirus (CMV) and human herpesvirus 6 by PCR. We have now repeated the infections with KSHV many times and find that all infected cultures survive longer than uninfected cells, which undergo apoptosis.

Thymidine incorporation was assayed to assess the proliferation rate of KSHV-infected ECs. The rates for infected HUVECs at 10 months post-infection (passage 50) were slightly higher than those for uninfected HUVECs at passage 10 (1,235 ± 334 c.p.m. versus 858 ± 6 c.p.m.), but were comparable to the known proliferation rates for early-passage HUVECs. Infected cultures responded to VEGF with a fourfold increase in proliferation rate, whereas uninfected cultures at this relatively late passage do not respond to VEGF. The thymidine incorporation for infected and uninfected HUVECs was 3,088 ± 164 c.p.m. versus 871 ± 230 c.p.m. (20 ng ml<sup>-1</sup> VEGF), and 4,791 ± 109 c.p.m. versus 769 ± 313 (40 ng ml<sup>-1</sup> VEGF). Therefore, although KSHV does not markedly increase the proliferation rate of ECs, it enables them to proliferate far beyond their natural lifespan in the presence of VEGF.

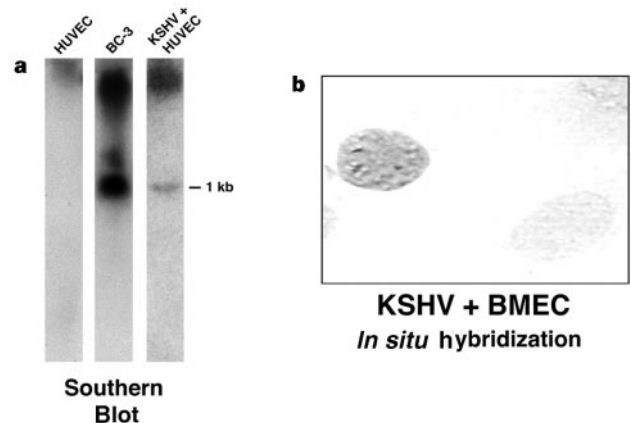
We investigated the induction of the KSHV lytic cycle by TPA treatment of infected ECs and evaluated the viral DNA copy number and expression of lytic transcripts using a quantitative TaqMan PCR and RT TaqMan-PCR, as well as straightforward PCR with reverse transcription. Induction was achieved, confirming that KSHV can establish a productive infection in human ECs (see Supplementary Information).

To investigate further the serial propagation of KSHV in HUVECs, Southern blotting was done on genomic DNA isolated from cells infected 4 months previously. After hybridization with a radiolabelled KSHV probe<sup>10</sup>, a band of the expected size indicated the presence of the KSHV genome (Fig. 2a). This result was confirmed with three additional viral probes (data not shown). We also used *in situ* hybridization to identify cells infected by KSHV<sup>3,11</sup> (Fig. 2b) and found that only ~5% of the cells contained KSHV DNA, which explains the faint band detectable by Southern blotting.

To determine whether infected ECs (BMECs and HUVECs)



**Figure 1** Infection of BMECs and HUVECs with KSHV. **a**, PCR to test for the presence of KSHV using KS330 primers, followed by hybridization with an internal probe. Primary infection (1° INF): BC-3 cells (lane 1) were induced with TPA, and KSHV viral particles (lane 2) were purified and concentrated from the supernatant. BMECs (lane 3) were infected with KSHV particles and two independently infected cultures four months after infection (lanes 4, 5). Secondary infection (2° INF): KSHV infected BMEC cells from primary infection (lane 4) were induced to enter lytic replication and purified KSHV particles were obtained (lane 6). These particles were used to infect early-passage HUVECs (lane 7) in triplicate (shown 40 days after infection in lanes 8–10). **b**, Growth in culture of KSHV-infected BMECs. Left, uninfected cells; middle, BMEC infected with KSHV particles (electron micrograph shown in insert labelled BC-3 KSHV); right, lytic replication, induced by decreasing the concentration of VEGF in the KSHV-infected BMEC culture.



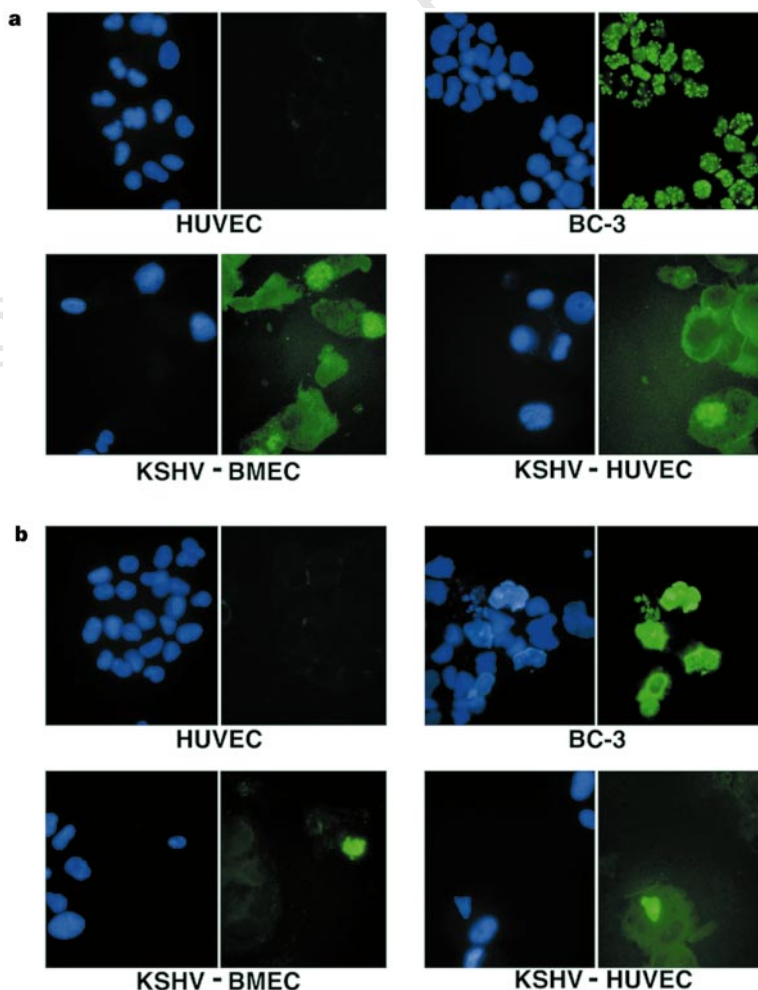
**Figure 2** Presence of KSHV DNA in infected endothelial cells. **a**, Southern blot analysis of DNA extracted from KSHV-infected and non-infected HUVEC. DNA was digested with *Bam*H1 and hybridized with a G-protein-coupled receptor radiolabelled probe. BC-3 cells were used as a positive control. A 72 h autoradiographic exposure for uninfected and infected HUVECs (left and right lanes), and a 16 h exposure for BC-3 DNA (middle lane) are shown. A positive band at 1 kb in the infected HUVECs indicates that the KSHV genome is present. **b**, TSA (tyramide signal-amplification system) *in situ* hybridization of KSHV-infected BMECs with a v-cyclin probe. The cell on the left was positive for KSHV DNA; the one on the right was negative.

expressed viral antigens, and to quantify the proportion of infected and inducible cells, we used an indirect immunofluorescence assay. Staining with serum from KS patients indicated that specific viral antigens were present in both KSHV-infected BMECs (8 months post-infection) and HUVECs (6 months post-infection) (Fig. 3a). Typical speckled nuclear staining was seen in latently infected cells, which comprised 1–5% of the total cell population, depending on the particular culture analysed. Staining with a polyclonal antibody raised against the major capsid protein (ORF 25) (not shown) and with monoclonal antibody against a DNA-replication protein (ORF 59)<sup>12</sup> (Fig. 3b) confirmed the ability of KSHV to replicate and express lytic viral antigens in infected ECs after TPA treatment. Less than 1% of all ECs in these cultures were positive after induction, and we estimate that ~10% of cells expressing latent antigens could be induced by TPA to undergo productive infection.

The phenotype of KSHV-infected ECs was similar to that of normal primary ECs (see Supplementary Information). However, serially passaged primary HUVECs and BMECs gradually lose expression of KDR (kinase domain receptor/Flk-1/VEGF receptor 2); KDR down-regulation is thought to be important in the senescence and death of ECs<sup>13</sup>. In contrast, all KSHV-infected HUVECs and BMECs continued to express significant amounts of KDR, even 11 months after infection. These results indicate that KSHV-infected ECs have the phenotype of primary ECs but have not downregulated expression of KDR after serial passage and so can still respond to VEGF.

As only a small subset of cells is infected by KSHV but uninfected cells also survive in these long-term cultures, we investigated whether a paracrine effect could account for this survival. We collected supernatants from infected HUVECs (10 months after infection) and from uninfected controls. These conditioned media were treated to remove herpesvirus particles and were tested for the presence of a variety of cytokines (see Supplementary Information). Uninfected HUVECs at passage 10 incubated with supernatants from KSHV-infected cultures acquired a spindle-shaped morphology (Fig. 4a); these cells were PCR-negative for KSHV. We tested these treated cultures for expression of KDR as upregulation of KDR constituted the most significant difference between infected and uninfected ECs: KDR expression was indeed upregulated, giving a twofold increase in fluorescence intensity for cells that had been treated with conditioned medium from infected cells as compared with cells treated with conditioned medium from uninfected cells (Fig. 4b). The paracrine induction of KDR expression therefore explains the proliferation of the uninfected cells in KSHV-positive cultures in response to VEGF in contrast to uninfected ECs, in which KDR is downregulated.

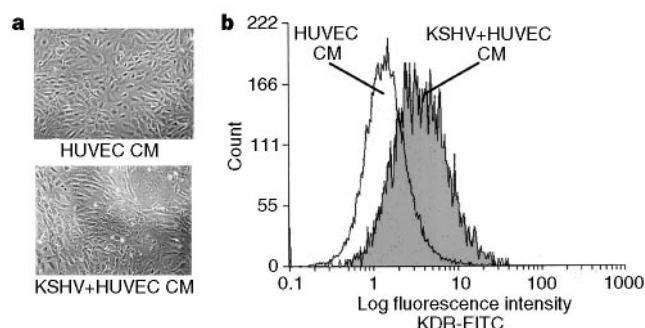
The normal limit to population doubling for a particular cell type, the 'Hayflick limit'<sup>14</sup>, must be surpassed in order for cells to exceed their usual lifespan. The enzyme telomerase is essential to maintain the telomere length of chromosomes and to enable cells to proliferate indefinitely<sup>15</sup>. We therefore measured the telomerase



**Figure 3** Detection of KSHV antigens in infected BMEC and HUVEC cells by immunofluorescence assay. **a**, Expression of latent antigens: KSHV-infected cells and controls were stained with serum from a KS patient. Positivity is demonstrated by speckled nuclear staining, as seen in the BC-3 control and scattered cells in the KSHV-infected BMEC and HUVEC cultures. **b**, Expression of

lytic antigens: the indicated cells were treated with TPA and stained with mAb 11D1 which recognizes a product of ORF 59, a DNA-replication protein. Positivity is demonstrated by diffuse nuclear staining. Cells were counterstained with DAPI (blue) to localize the nuclei, shown on the left of each photomicrograph.



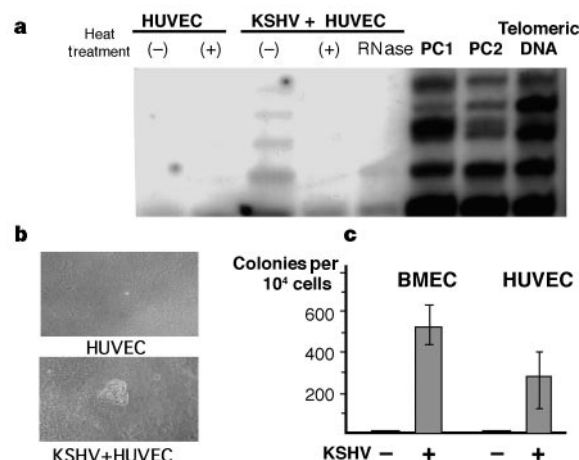


**Figure 4** Paracrine effects in uninfected HUVEC cells by conditioned medium from KSHV-positive endothelial cell cultures. **a**, Morphology of HUVEC cells, treated with filtered conditioned media (CM) from uninfected HUVECs (top) or KSHV-infected HUVECs (bottom). **b**, FACS analysis for expression of KDR (VEGFR-2) in passage-10 HUVEC cells treated with filtered conditioned medium from uninfected HUVECs (mean log fluorescence, 1.44) or KSHV-infected HUVECs (mean log fluorescence, 3.68; grey area).

activity in our cultures by using a telomeric-repeat amplification protocol (TRAP)<sup>16</sup> and found that only the cell cultures infected with KSHV had telomerase activity. Telomerase was undetectable in uninfected cells, with and without exposure to a supernatant from an infected culture (Fig. 5a). This upregulation of telomerase activity in cultures infected with KSHV provides additional evidence for a role for KSHV in the long-term survival of ECs.

ECs need the extracellular matrix for proliferation and survival: loss of this anchorage dependence is a feature of cellular transformation. We tested our KSHV-infected ECs for anchorage-independent growth by determining their ability to form colonies in soft agar. The experiments shown in Fig. 5b, c were done using BMECs and HUVECs at 12 and 10 months post-infection respectively, but results were comparable for several cellular stocks at 6 and 9 months post-infection (BMECs), as well as at 4 and 7 months post-infection (HUVECs). Small colonies were evident after 5 days of culture at 37°C and were clearly visible after 12 d. After two weeks, these colonies measured on average 250 µm (BMECs) and 30 µm (HUVECs). In contrast, uninfected cell cultures did not form any colonies, which excludes the possibility that colony formation is due to contaminating non-ECs, such as bone-marrow progenitor cells. An average of 400 (range 120 to 624) colonies was obtained from 10<sup>4</sup> cells plated (Fig. 5c). Therefore ~4% (range, 1–6%) of plated cells result in colony formation, which is comparable to our estimate of the proportion of KSHV-positive cells in these cultures. To determine whether it was only the KSHV-positive cells that were able to grow in soft agar, we collected 50 individual colonies (30 BMEC and 20 HUVEC), as well as portions of agar without colonies, by microdissection and analysed them by PCR for KSHV. All colonies contained viral DNA except one, which also failed to amplify with a control set of primers (data not shown). We did not detect any KSHV in agar that contained no colonies, ruling out the possibility that diffusion of KSHV could be responsible for positivity in all the colonies. This result indicates that KSHV is necessary for anchorage-independent growth.

We have shown KSHV can induce transformation of primary ECs, as defined by extended survival, acquisition of telomerase activity and anchorage-independent growth. We have found that the mechanism for this apparent immortalization is different from that reported for other transforming viruses, because KSHV is present in only 1–6% of cells in each infected culture. Although KSHV does not seem to provide a proliferative advantage to infected or uninfected cells, it uses a paracrine mechanism to enable uninfected cells in the same culture to proliferate beyond their natural lifespan; this proliferation is due to the paracrine upregulation of KDR causing these cells to respond to VEGF. KSHV



**Figure 5** Transformation of KSHV-infected HUVECs. **a**, Telomerase activity in KSHV-infected HUVECs is demonstrated by the presence of a DNA ladder in the TRAP assay, the formation of which is inhibited by heat or RNase treatment. Uninfected HUVECs had no telomerase activity. Positive controls include two telomerase-positive cell lines (PC1 and PC2). **b**, Infected and non-infected HUVECs cultured in soft agar. After 12 days, only isolated cells were observed in the non-infected culture (top). In contrast, KSHV-infected cells formed distinct colonies (bottom). **c**, Number of colonies obtained after plating 10<sup>4</sup> cells per plate in triplicate.

also induces anchorage-independent growth, a feature of transformation. In contrast to the requirements for long-term survival, KSHV is needed for growth in soft agar, where all colonies contain the viral genome. Our results are consistent with a transforming role for this virus in the pathogenesis of Kaposi's sarcoma. □

## Methods

**Cell culture, viral purification and infection.** Primary bone-marrow microvascular ECs (BMECs) and human umbilica vein ECs (HUVECs) were infected with purified KSHV particles (see Supplementary Information). To determine the effect of conditioned media, KSHV-infected and uninfected HUVECs and BMECs were cultured for 72 h in fresh medium, after which these conditioned media were collected. KSHV particles were removed by filtration using a syringe filter (0.1-µm pore size) or by centrifugation at 15,000g for 2 h. Uninfected HUVECs were incubated with conditioned medium for 48 h and tested for expression of KDR by flow cytometric analysis.

**Southern blots.** DNA extraction and Southern blot analysis followed standard procedures<sup>5</sup>. A GPCR probe (ORF 74) was obtained by PCR amplification as described<sup>10</sup>.

**PCR and RT-PCR.** KS330 primers were used for PCR as described<sup>1</sup>. To detect KSHV in colonies growing in soft agar, we used a microcapillary tube to withdraw cells in a colony under a dissecting microscope; the DNA was extracted and amplified using Generelease (Bio-Ventures). Amplified products were electrophoresed, transferred onto nylon membranes and hybridized with a <sup>32</sup>P-radiolabelled internal probe. Total RNA was treated with 2 units of RNase-free DNase (Promega) and reverse-transcribed. We used primer sets for K2 (v-IL6), ORF 73 (LANA) and ORF 74 (GPCR) for PCR<sup>10</sup>.

**TaqMan PCR.** TaqMan PCR has been described<sup>17</sup>; primers and the TaqMan probe were designed using the 'nearest neighbour method'<sup>18</sup>. Specific products were detected using a Perkin-Elmer 7700 sequence detector for real-time PCR. Primer sequences and the specific conditions used for each primer set are available as Supplementary Information.

**In situ hybridization to detect KSHV.** Hybridization of spotted, fixed cells was done after denaturation with a PCR-generated v-cyclin probe, double-labelled with digoxigenin and biotin. Controls used included DNase- and RNase-treated cells, an irrelevant probe (HPV16), negative control tissue (uninfected ECs) and positive control cells (BC-3 cells). An indirect tyramide signal-amplification system was used (TSA indirect) (NEN Life Sciences) (see Supplementary Information).

**Immunofluorescence assay.** For latent KSHV antigens, cytospin preparations were fixed with methanol/acetone, and incubated with KS sera at a 1:40 dilution in master block as described<sup>19</sup>, followed by incubation with an FITC-conjugated F(ab')<sub>2</sub> fragment of rabbit anti-human IgG (DAKO). For lytic antigens, cytospin preparations of TPA-induced cells were fixed and incubated either overnight at 4 °C with mAb 11D1 (ref. 12; gift from B. Chandran), which is specific for the DNA-replication protein ORF 59, at a dilution of 1:10, or for 1 h at room temperature with a polyclonal rabbit antibody raised against the main capsid protein ORF 25 (gift from S. J. Gao) and diluted 1:300, followed by 40 min incubation with an FITC-conjugated goat F(ab')<sub>2</sub> fragment to either mouse IgG (ICN Biomedicals) or rabbit IgG (Jackson ImmunoResearch). Cell nuclei were counterstained with DAPI (4'-6'-diamino-2'-phenylindole dihydrochloride; Boehringer Mannheim), and examined in a fluorescence microscope (Axioplan 2; Zeiss).

**Cell-proliferation assay.** Infected and uninfected HUVECs were washed and cultured at 10<sup>4</sup> cells per 200 µl in serum-free medium (X-vivo 20) containing no VEGF, or 20 ng ml<sup>-1</sup> or 40 ng ml<sup>-1</sup> VEGF for 24 h. Cells were then pulsed with <sup>3</sup>H-thymidine, maintained in culture for 36 h, collected and counted in a β-scintillation counter. Values represent the mean ± standard deviation c.p.m. <sup>3</sup>H-thymidine incorporated for triplicate cultures.

**Telomerase assay.** The radioactive TRAP assay for telomerase activity has been described<sup>16,20</sup>.

**Anchorage-independent growth.** 2 × M 199 medium (BioWhittaker) with 2 × ECGS (endothelial growth supplement, PerImmune, Inc.) and 40% fetal calf serum (FCS) were mixed with an equal amount of 0.8% melted agar and 1 ml was poured into each 35-mm plate to form a bottom layer; then 2 × M 199 with 2 × ECGF and 40% FCS were mixed with an equal amount of 0.4% melted agar and 1 ml was added to serial concentrations of cells (10<sup>3</sup>, 10<sup>4</sup>, 10<sup>5</sup>) a. 1 poured over the bottom layer to form a top layer.

**Cytokine and chemokine assays.** Assays were done by ELISA in a commercial laboratory (Cytokine Core Laboratory).

**Flow cytometer analysis.** FITC conjugated mAbs (PECAM, E-selectin, VCAM, ICAM) were used as described<sup>8</sup>. KDR expression was quantified by a FITC-conjugated mAb to KDR (clone 6.64, VEGF-binding domain) (gift from ImClone Systems).

Received 30 March; accepted 11 June 1998.

- Chang, Y. *et al.* Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**, 1865–1869 (1994).
- Moore, P. S. & Chang, Y. Detection of herpesvirus-like DNA sequences in Kaposi's sarcoma in patients with and those without HIV infection. *N. Engl. J. Med.* **332**, 1181–1185 (1995).
- Boshoff, C. *et al.* Kaposi's sarcoma-associated herpesvirus infects endothelial and spindle cells. *Nature Med.* **1**, 1274–1278 (1995).
- Reed, J. A. *et al.* Demonstration of Kaposi's sarcoma-associated herpesvirus cyclin D homolog in cutaneous Kaposi's sarcoma by colorimetric *in situ* hybridization using a catalyzed signal amplification system. *Blood* (in the press).
- Arvanitakis, L. *et al.* Establishment and characterization of a primary effusion (body cavity-based) lymphoma cell line (BC-3) harboring Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8) in the absence of Epstein-Barr virus. *Blood* **88**, 2648–2654 (1996).
- Renne, R. *et al.* Lytic growth of Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in culture. *Nature Med.* **2**, 342–346 (1996).
- Miller, G. *et al.* Selective switch between latency and lytic replication of Kaposi's sarcoma herpesvirus and Epstein-Barr virus in dually infected body cavity lymphoma cells. *J. Virol.* **71**, 314–324 (1997).
- Rafii, S. *et al.* Human bone marrow microvascular endothelial support long term proliferation and differentiation of myeloid and megakaryocytic progenitors. *Blood* **86**, 3353–3363 (1995).
- Watanabe, Y. & Dvorak, H. F. Vascular permeability factor/vascular endothelial growth factor inhibits anchorage-disruption-induced apoptosis in microvessel endothelial cells by inducing scaffold formation. *Exp. Cell Res.* **233**, 340–349 (1997).
- Cesarman, E. *et al.* Kaposi's sarcoma associated herpesvirus (KSHV/HHV-8) contains G protein-coupled receptor and cyclin D homologs which are expressed in Kaposi's sarcoma and malignant lymphoma. *J. Virol.* **70**, 8218–8223 (1996).
- O'Leary, J. *PCR In-situ Hybridization*. 1–53–86 (IRL, Oxford University Press, Oxford, 1997).
- Ruby Chan, S., Bloomer, C. & Chandran, B. Identification and characterization of human herpesvirus-8 lytic cycle associated ORF 59 protein and the encoding cDNA by monoclonal antibody. *Virology* **240**, 118–126 (1998).
- Hewett, P. W. & Murray, J. C. Coexpression of flt-1, flt-4 and KDR in freshly isolated and cultured human endothelial cells. *Biochem. Biophys. Res. Commun.* **221**, 697–702 (1996).
- Rienke, U., Burlington, H., Cronkite, E. P. & Laissue, J. Hayflick's hypothesis: an approach to *in vivo* testing. *Fed. Proc.* **34**, 71–75 (1975).
- Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011–2015 (1994).
- Mieczyslaw, A. P. *et al.* Detection of telomerase activity in human cells and tumors by a telomeric repeat amplification protocol (TRAP). *Meth. Cell Sci.* **17**, 1–15 (1995).
- Kennedy, M. M. *et al.* HHV8 and Kaposi's sarcoma: a time cohort study. *Mol. Pathol.* **50**, 96–100 (1997).
- Rhoads, R. E. Optimization of the annealing temperature for DNA amplification *in vitro*. *Nucleic Acids Res.* **18**, 6409–6412 (1990).
- Kedes, D. H. *et al.* The seroepidemiology of human herpesvirus 8 (Kaposi's sarcoma-associated herpesvirus): Distribution of infection in KS risk groups and evidence for sexual transmission. *Nature Med.* **2**, 918–924 (1996).

20. Kim, N. W. & Wu, F. Advances in quantification and characterization of telomerase activity by the telomeric repeat amplification protocol (TRAP). *Nucleic Acids Res.* **25**, 2595–2597 (1997).

**Supplementary information** is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank S. Chen-Kiang, D. Knowles, M. C. Gershengorn and E. Mesri for their suggestions; B. Chandran and S.-J. Gao for antibodies against lytic antigens; P. Pellet for positive controls CMB and HHV-6; L. Friedman, S. Nikolovska, B. Ferris, G. Lam, A. J. Naiyer, S. Diana, I. Silva and V. Uhlmann for technical help; and S. Picton and K. Lohman at Perkin Elmer Applied Biosystems.

Correspondence and requests for materials should be addressed to E.C. (e-mail: [ecesar@med.cornell.edu](mailto:ecesar@med.cornell.edu)).

## Perinuclear localization of chromatin facilitates transcriptional silencing

Erik D. Andrulis, Aaron M. Neiman, David C. Zappulla & Rolf Sternglanz

Department of Biochemistry and Cell Biology, State University of New York at Stony Brook, Stony Brook, New York 11794-5215, USA

**Transcriptional silencing in *Saccharomyces cerevisiae* at the *HM* mating-type loci and telomeres occurs through the formation of a heterochromatin-like structure. *HM* silencing is regulated by *cis*-acting elements, termed silencers, and by *trans*-acting factors that bind to the silencers. These factors attract the four SIR (silent information regulator) proteins, three of which (SIR2–4) spread from the silencers to alter chromatin, hence silencing nearby genes<sup>1–4</sup>. We show here that an *HMR* locus with a defective silencer can be silenced by anchoring the locus to the nuclear periphery. This was accomplished by fusing integral membrane proteins to the GAL4 DNA-binding domain and overproducing the hybrid proteins, causing them to accumulate in the endoplasmic reticulum and the nuclear membrane. We expressed the hybrid proteins in a strain carrying an *HMR* silencer with GAL4-binding sites (UAS<sub>G</sub>) replacing silencer elements, causing the silencer to become anchored to the nuclear periphery and leading to silencing of a nearby reporter gene. This silencing required the hybrids of the GAL4 DNA-binding domain with membrane proteins, the UAS<sub>G</sub> sites and the SIR proteins. Our results indicate that perinuclear localization helps to establish transcriptionally silent chromatin.**

Yeast and *Drosophila* telomeres are probably located at the nuclear periphery<sup>5–7</sup>, where some non-telomeric heterochromatin is also found<sup>8</sup>; however, it is unclear whether this localization contributes to the formation of heterochromatin or is a consequence of it. We tested whether genes artificially drawn to the yeast nuclear envelope could become transcriptionally silenced by using strains with an *HMR* locus in which UAS<sub>G</sub> sites replaced some or all of the usual elements at the *HMR-E* silencer. This silencer normally contains three elements, A, E and B, that are binding sites for the proteins ORC, RAP1 and ABF1, respectively<sup>9</sup>. Deletion of any one of these elements leads to little or no derepression of *HMR*, whereas deletion of any two elements leads to full derepression<sup>9</sup>. If these silencer elements are replaced by UAS<sub>G</sub> sites, silencing is lost, but it can be restored by the introduction of any one of several different GAL4 DNA-binding domain (G<sub>BD</sub>) hybrids, including those with SIR1, SIR3, SIR4, RAP1 and ORC1 (refs 2, 10–12): we refer to this as targeted silencing. We used the same strategy to assess the role of perinuclear localization in silencing. As overexpression of Golgi proteins leads to their accumulation in the endoplasmic reticulum (ER) and the nuclear membrane<sup>13,14</sup>, we investigated whether overexpressed G<sub>BD</sub>-Golgi protein hybrids could cause targeted silencing.

First, we tested YIP1, an essential protein with at least three predicted membrane-spanning domains. YIP1 is a Golgi protein

involved in ER-to-Golgi and/or early Golgi membrane transport (D. Gallwitz, personal communication). Indirect immunofluorescence showed that when  $G_{BD}$ -YIP1 was overproduced, nuclear rim staining appeared (Fig. 1) that was consistent with accumulation in the ER and the nuclear membrane. In contrast,  $G_{BD}$  alone was seen throughout the nucleoplasm, coincident with DAPI staining.

We tested the  $G_{BD}$ -YIP1 hybrid in three targeted silencing strains<sup>10</sup>, which differed only in the specific silencer elements deleted from the *HMR-E* silencer. A cell-spotting assay was used to quantify the extent of silencing of an *hmr::TRP1* reporter gene. Lack of growth on a medium without tryptophan demonstrates silencing in this assay. The  $G_{BD}$ -YIP1 hybrid gave silencing in strain YSB35 (Fig. 2a, row 1), which has the RAP1- and ABF1-binding sites deleted from the *HMR-E* silencer. This hybrid also gave silencing in strain YSB2, which has the ORC- and RAP1-binding sites deleted (Fig. 2b, row 1), although there was less silencing than in YSB35 (compare rows 1 in Fig. 2a, b). This had been seen previously for  $G_{BD}$ -SIR1; it too gave better targeted silencing in YSB35 than in YSB2 (ref. 10).

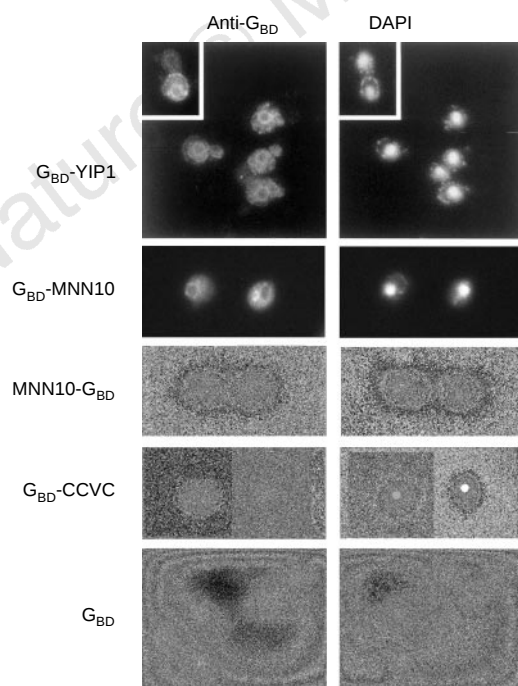
Figure 2 also shows the effect of *sir* mutations in these strains. Silencing by  $G_{BD}$ -YIP1 was abolished in *sir2* and *sir3* derivatives of strain YSB35 and greatly weakened in a *sir1* mutant (Fig. 2a). Silencing by  $G_{BD}$ -YIP1 in strain YSB2 also depended on SIR2, SIR3 and SIR4, but was barely affected by a *sir1* mutation (Fig. 2b, and data not shown). The silencing also depended on the presence of the  $UAS_G$  sites (Fig. 2d), and thus had all the hallmarks of targeted silencing (demonstrated previously by proteins directly involved in silencing)<sup>2,10-12</sup>.

There was no silencing for  $G_{BD}$ -YIP1 in strain YSB41, which has all three elements deleted from the *HMR-E* silencer (Fig. 2c), in contrast to  $G_{BD}$ -SIR1, which gave some silencing in this strain<sup>2</sup> (Fig. 2c). Thus, targeted silencing by this membrane protein required the presence of at least one natural silencer element.

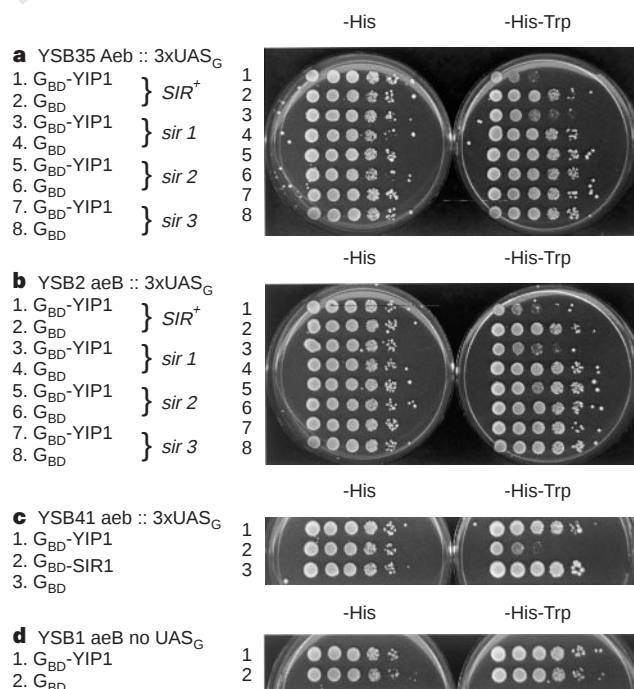
To test the generality of these results with  $G_{BD}$ -YIP1, we tested four other membrane proteins. We identified two as YIP1-interact-

ing proteins in a two-hybrid screen, YIF1 and YIP3. YIP3 is also involved in ER-to-Golgi membrane transport (D. Gallwitz, personal communication). It has two predicted membrane-spanning domains, whereas YIF1 has three. The other two proteins tested were MNN10, a Golgi mannosyltransferase<sup>15</sup>, and STT3, a resident ER protein that is part of the oligosaccharyltransferase complex<sup>16,17</sup>. In the case of MNN10, overexpression leads to accumulation of the protein in the ER membrane (N. Dean, personal communication). As the N termini of both membrane proteins are likely to be cytoplasmic rather than luminal<sup>15,16</sup>, overexpression of  $G_{BD}$  hybrids of these proteins should cause their N termini to reside in the cytoplasm and/or nucleoplasm. All four membrane-protein hybrids caused repression of the *hmr::TRP1* reporter gene in strain YSB35 (Fig. 3a). Quantitative measurements showed that the degree of targeted silencing for the YIF1 hybrid was as good as, or better than, that of SIR1. The MNN10, YIP1 and YIP3 hybrids each gave less silencing than SIR1 and YIF1 (see Supplementary Information); the STT3 hybrid gave the least (Fig. 3). The silencing by these membrane-protein hybrids had all the features of targeted silencing, in that it required  $UAS_G$  sites and depended on SIR2, 3 and 4, as did  $G_{BD}$ -YIP1 (Fig. 2). All of the proteins gave better silencing in strain YSB35 than in YSB2, and none in YSB41, as did  $G_{BD}$ -YIP1 (Fig. 2).

We confirmed that this silencing was not due simply to overproduction of membrane proteins in the following experiment. As the N terminus of MNN10 is probably cytoplasmic, the  $G_{BD}$  domain



**Figure 1**  $G_{BD}$  hybrids localized by indirect immunofluorescence. Cells were stained with DAPI (diamidino-2-phenylindole) to visualize the nuclei and with an anti- $G_{BD}$  antibody to visualize  $G_{BD}$  and  $G_{BD}$  hybrids. DAPI staining shows the location of the nuclei and the mitochondria. Note that  $G_{BD}$  hybrids are observed at the nuclear rim and in some cases (inset) under the plasma membrane, consistent with ER localization<sup>26</sup>.



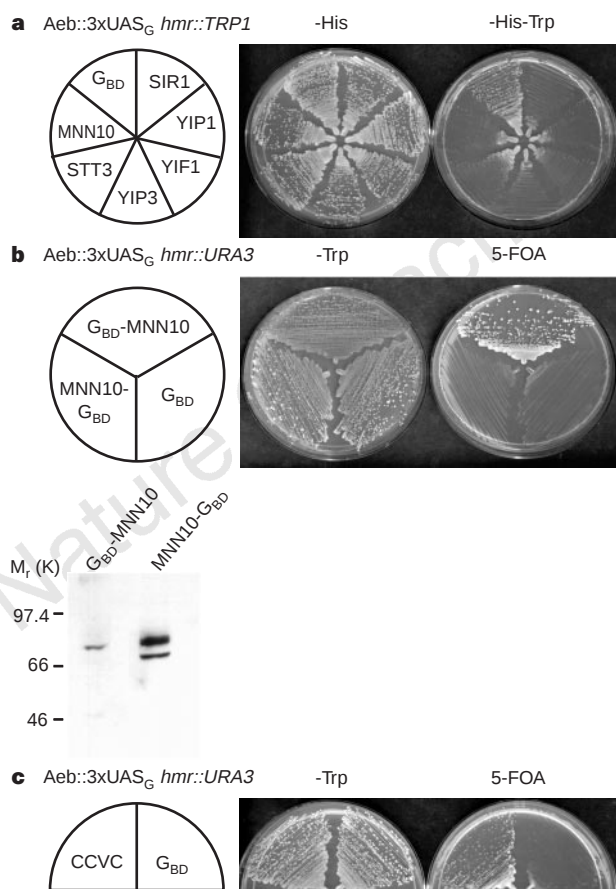
**Figure 2** Targeted silencing by  $G_{BD}$ -YIP1. **a**,  $G_{BD}$ -YIP1 or  $G_{BD}$  was expressed in strain YSB35 (rows 1 and 2) or its *sir1* (rows 3 and 4), *sir2* (rows 5 and 6) or *sir3* (rows 7 and 8) derivatives. These strains have the E and B elements deleted from the *HMR-E* silencer and replaced by three  $UAS_G$  sites (designated Aeb)<sup>10</sup>. **b**,  $G_{BD}$ -YIP1 or  $G_{BD}$  was expressed in strain YSB2 (rows 1 and 2) or its *sir1* (rows 3 and 4), *sir2* (rows 5 and 6) or *sir3* (rows 7 and 8) derivatives. These strains have the A and E elements deleted from the *HMR-E* silencer and replaced with three  $UAS_G$  sites (designated aeb)<sup>10</sup>. **c**,  $G_{BD}$ -YIP1,  $G_{BD}$ -SIR1 or  $G_{BD}$  was expressed in strain YSB41. This strain has all three elements deleted from the *HMR-E* silencer and replaced with three  $UAS_G$  sites (designated aeb)<sup>10</sup>. **d**,  $G_{BD}$ -YIP1 or  $G_{BD}$  was expressed in strain YSB1. This strain has a similar silencer to the one in YSB2 (aeb, **b**), but has no  $UAS_G$  sites. In all cases, cell cultures were serially diluted 10-fold five times and 5  $\mu$ l of each dilution spotted from left (no dilution) to right ( $10^5$ -fold dilution). Growth was assessed after two days at 30°C. Targeted silencing is seen as a lack of growth on -His-Trp medium (for example, rows 1 and 3 in **a**, **b**, and row 2c).

of the  $G_{BD}$ -MNN10 hybrid, when it is overexpressed, should be in the nucleoplasm. As a control, we therefore constructed an MNN10 hybrid with  $G_{BD}$  at the C terminus (MNN10- $G_{BD}$ ). This protein would be expected to have  $G_{BD}$  in the lumen of the ER rather than in the nucleus, and thus would not be expected to give targeted silencing, which we found to be the case (Fig. 3b). This hybrid protein was expressed more than  $G_{BD}$ -MNN10 (Fig. 3b), and immunofluorescence experiments localized both MNN10 hybrids to the ER/nuclear membrane (Fig. 1).

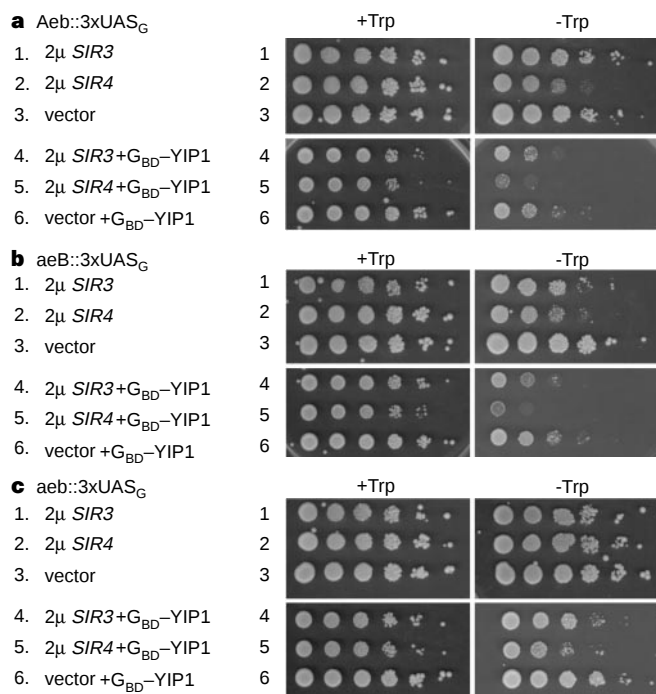
We conclude that the  $G_{BD}$ -membrane protein hybrids, when overproduced, reside in the inner nuclear membrane and draw  $UAS_G$  sites to the nuclear periphery. We confirmed this by using a screen (to be described elsewhere) in which we looked for plasmids encoding  $G_{BD}$  hybrid proteins that would give targeted silencing in strain YEA76 (Aeb::3xUAS<sub>G</sub> *hmr::URA3*). Among the plasmids found was one coding for  $G_{BD}$  fused to a tetrapeptide, Cys-Cys-Val-Cys (CCVC), followed by a stop codon; this tetrapeptide led to targeted silencing of a *URA3* reporter gene (Fig. 3c). The silencing did not occur in the same strain containing a *sir3* mutation. Furthermore, a frameshift mutation introduced at the junction between  $G_{BD}$  and the tetrapeptide abolished silencing (data not shown). As proteins with the sequence Cys-X-Cys

(where X is any amino acid) at the C terminus are geranylgeranylated<sup>18,19</sup>, we presume that the  $G_{BD}$ -CCVC hybrid was similarly modified; this modification then anchors the protein in the nuclear membrane (Fig. 1), where it facilitates targeted silencing in the same way as did the authentic membrane proteins. Thus, silencing is a consequence of anchorage to the periphery, rather than the attachment of a protein transmembrane domain to  $G_{BD}$  *per se*.

The membrane protein hybrids described here are unlikely to participate directly in silencing. Our data suggest that a membrane protein, when fused to  $G_{BD}$  and overexpressed, can bind to  $UAS_G$  sites on the chromosome and recruit nearby genes to the nuclear periphery. Why does such localization facilitate silencing? It has been suggested that the concentration of limiting SIR proteins is higher at the nuclear periphery<sup>5,12,20</sup>, so recruitment to the periphery may rescue the silencing defect by increasing the local concentration of SIR proteins. One prediction of this model is that overproduction of SIR proteins may restore some silencing to partially defective silencers. Such overproduction can improve silencing in several contexts<sup>20-22</sup>. We found that overexpression of SIR4, and to a lesser extent SIR3, in strains YSB35 and YSB2 (Aeb and aeb silencers, respectively) caused some silencing (Fig. 4a, b). Furthermore, a combination of SIR3 or SIR4 overexpression, together with  $G_{BD}$ -YIP1, gave even greater silencing than did  $G_{BD}$ -YIP1 alone (Fig. 4). In strain YSB41, with all three silencer elements deleted from *HMR-E*, overexpression of SIR4 or SIR3 alone did not cause silencing (Fig. 4c), and neither did the membrane-protein hybrid  $G_{BD}$ -YIP1 (Figs 2c, 4c). However, the combination of SIR4 (or SIR3) overexpression plus  $G_{BD}$ -YIP1 did give silencing in this strain (Fig. 4c). These results support the idea that concentrations of SIR proteins are greater at the nuclear periphery, and are in agreement with previous immunofluorescence results<sup>5</sup>. We conclude that peri-



**Figure 3** Several membrane protein hybrids give targeted silencing. **a**, Strain YSB35 expressing  $G_{BD}$ ,  $G_{BD}$ -SIR1,  $G_{BD}$ -YIP1,  $G_{BD}$ -YIF1,  $G_{BD}$ -YIP3,  $G_{BD}$ -STT3 or  $G_{BD}$ -MNN10 was streaked either on -His medium (control) or on -His-Trp medium and growth assessed after two days at 30°. Lack of growth on -His-Trp is indicative of silencing. **b**, Strain YEA76 (Aeb::3xUAS<sub>G</sub> *hmr::URA3*) expressing  $G_{BD}$ -MNN10, MNN10- $G_{BD}$ , or  $G_{BD}$  was streaked on -Trp medium (control) or on 5-fluoro-orotic acid (5-FOA)-containing medium. Growth on the latter medium is indicative of silencing. The lower panel shows a western blot of  $G_{BD}$ -MNN10 and MNN10- $G_{BD}$  expressed in this strain. **c**, Strain YEA76 expressing  $G_{BD}$ -CCVC or  $G_{BD}$ . Silencing was assessed by growth on 5-FOA-containing medium as in **b**.



**Figure 4** Overexpression of SIR3 and SIR4 improves silencing in strains with defective silencers. **a**, Strain YSB35; **b**, strain YSB2; **c**, strain YSB41. These strains are the same as the ones used for Fig. 2; they differ only in the nature of their silencers, as indicated. Strains with the indicated plasmids were serially diluted on control medium (-Ura in the case of one plasmid, or -Ura-His in the case of two plasmids; indicated as +Trp) or on medium lacking tryptophan (-Trp). As in Fig. 2, lack of growth on -Trp medium reflects silencing.



nuclear localization is important in establishing transcriptionally silent chromatin. □

## Methods

**Strains and plasmids.** Strains YSB35, YSB2, YSB41 and YSB2 were used to measure targeted silencing as described<sup>10</sup>. All five membrane proteins were cloned in-frame into the overexpression vector pMA424 (ref. 23) to create G<sub>BD</sub> hybrids. The entire coding sequence was used for YIP1 (pEDA73) and MNN10 (pEDA96). The YIP3 hybrid encompassed amino acids 39–182 (pEDA85); the YIF1 hybrid, amino acids 55–314 (pEDA76); and the STT3 hybrid, amino acids 45–720 (pEDA93). G<sub>BD</sub>–MNN10 (pEDA109) was constructed by digesting pEDA96 with *EcoRI* and *SalI* and subcloning the MNN10 fragment into pGBT9C. To make MNN10–G<sub>BD</sub>, a PCR fragment with the MNN10 coding region was cloned into the *BamHI* site of the C-terminal G<sub>BD</sub> fusion vector D134 (gift from R. Brazas). The SIR3 and SIR4 overexpression plasmids were pJRI04 and pHR643, respectively (from J. Rine's laboratory).

**Indirect immunofluorescence and western blot analysis.** For immunofluorescence, cells were grown to mid-log phase, fixed with formaldehyde and prepared as described<sup>24</sup>. The primary antibody used was one directed against G<sub>BD</sub> (Upstate Biotechnology). For western blots, strain YSB35 transformants were grown to an absorbance at 600 nm of 1.0 and lysed by vortexing with glass beads. Western blotting was done as described<sup>25</sup> using the same anti-G<sub>BD</sub> primary antibody to detect G<sub>BD</sub> hybrids.

Received 23 March; accepted 16 June 1998.

- Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S. M. & Grunstein, M. Histone H3 and H4 N-termini interact with the silent information regulators Sir3 and Sir4: A model for the formation of heterochromatin in yeast. *Cell* **80**, 583–592 (1995).
- Triolo, T. & Sternglanz, R. Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. *Nature* **381**, 251–253 (1996).
- Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repression by SIR3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
- Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. SIR2 and SIR4 interactions in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11**, 83–93 (1997).
- Gotta, M. *et al.* The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type *Saccharomyces cerevisiae*. *J. Cell Biol.* **134**, 1349–1363 (1996).
- Hiraoka, Y., Agard, D. A. & Sedat, J. W. Temporal and spatial coordination of chromosome movement, spindle formation, and nuclear envelope breakdown during prometaphase in *Drosophila melanogaster* embryos. *J. Cell Biol.* **11**, 2815–2828 (1990).
- Funabiki, H., Hagan, L., Usawa, S. & Yanagida, M. Cell cycle dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J. Cell Biol.* **121**, 961–976 (1993).
- Marshall, W. F., Dernburg, A. F., Harmon, B., Agard, D. A. & Sedat, J. W. Specific interactions of chromatin with the nuclear envelope: Positional determination within the nucleus in *Drosophila melanogaster*. *Mol. Biol. Cell* **7**, 825–842 (1996).
- Brand, A. H., Breeden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. Characterization of a "silencer" in yeast: A DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* **41**, 41–48 (1985).
- Chien, C.-T., Buck, S., Sternglanz, R. & Shore, D. Targeting of SIR1 protein establishes transcriptional silencing at *HM* loci and telomeres in yeast. *Cell* **75**, 531–541 (1993).
- Buck, S. W. & Shore, D. Action of a RAP1 carboxy-terminal silencing domain reveals an underlying competition between *HM* and telomeres in yeast. *Genes Dev.* **9**, 370–384 (1995).
- Marcand, S., Buck, S. W., Moretti, P., Gilson, E. & Shore, D. Silencing of genes at nontelomeric sites in yeast is controlled by sequestration of silencing factors at telomeres by Rap1 protein. *Genes Dev.* **10**, 1297–1309 (1996).
- Munro, S. Sequences within and adjacent to the transmembrane segment of the  $\alpha$ -2,6-sialyltransferase specify Golgi retention. *EMBO J.* **10**, 3577–3588 (1991).
- Machamer, C. E. *et al.* Retention of a cis Golgi protein requires polar residues on one face of a predicted  $\alpha$ -helix in the transmembrane domain. *Mol. Biol. Cell* **4**, 695–704 (1993).
- Dean, N. & Poster, J. Molecular and phenotypic analysis of the *S. cerevisiae* MNN10 gene identifies a family of related glycosyltransferases. *Glycobiology* **6**, 73–81 (1996).
- Zufferey, R. *et al.* STT3, a highly conserved protein required for yeast oligosaccharyl transferase activity *in vivo*. *EMBO J.* **14**, 4949–4960 (1995).
- Karaoglu, D., Kelleher, D. J. & Gilmore, R. The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. *J. Biol. Chem.* **272**, 32513–32520 (1997).
- Khosravi-Far, R. *et al.* Isoprenoid modification of rab proteins terminating in CC or CXC motifs. *Proc. Natl Acad. Sci. USA* **88**, 6264–6268 (1991).
- Kinsella, B. T. & Maltese, W. A. *rab* GTP-binding proteins with three different carboxyl-terminal cysteine motifs are modified *in vivo* by 20-carbon isoprenoids. *J. Biol. Chem.* **267**, 3940–3945 (1992).
- Maillet, L. *et al.* Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes Dev.* **10**, 1796–1811 (1996).
- Stone, E. M., Swanson, M. J., Romeo, A. M., Hicks, J. B. & Sternglanz, R. The SIR1 gene of *Saccharomyces cerevisiae* and its role as an extragenic suppressor of several mating-defective mutants. *Mol. Cell Biol.* **11**, 2253–2262 (1991).
- Renauld, H. *et al.* Silent domains are assembled continuously from the telomere and are defined by promoter distance and strength, and SIR3 dosage. *Genes Dev.* **7**, 1133–1145 (1993).
- Ma, J. & Ptashne, M. A new class of yeast transcriptional activators. *Cell* **51**, 113–119 (1987).
- Pringle, J. R., Adams, A. E. M., Drubin, G. & Haarer, B. K. Guide to yeast genetics and molecular biology. *Meth. Enzymol.* **194**, 565–602 (1991).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (CSHL Press, Cold Spring Harbor, NY, 1988).
- Kaiser, C. A. & Schekman, R. Distinct sets of SEC genes govern transport vesicle formation and fusion early in the secretory pathway. *Cell* **61**, 723–733 (1990).

**Supplementary information** is available on Nature's World Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

**Acknowledgements.** We thank D. Gallwitz for communicating unpublished results; N. Dean, H. Park, S. Marcand, D. Shore and R. Brazas for plasmids and strains; and N. Dean, A. Mook and C. Dingwall for advice and comments. This work was supported by a grant from the NIH.

Correspondence and requests for materials should be addressed to R.S. (e-mail: [rolf@life.bio.sunysb.edu](mailto:rolf@life.bio.sunysb.edu)).

# Crystal structure of a small heat-shock protein

Kyeong Kyu Kim, Rosalind Kim & Sung-Hou Kim

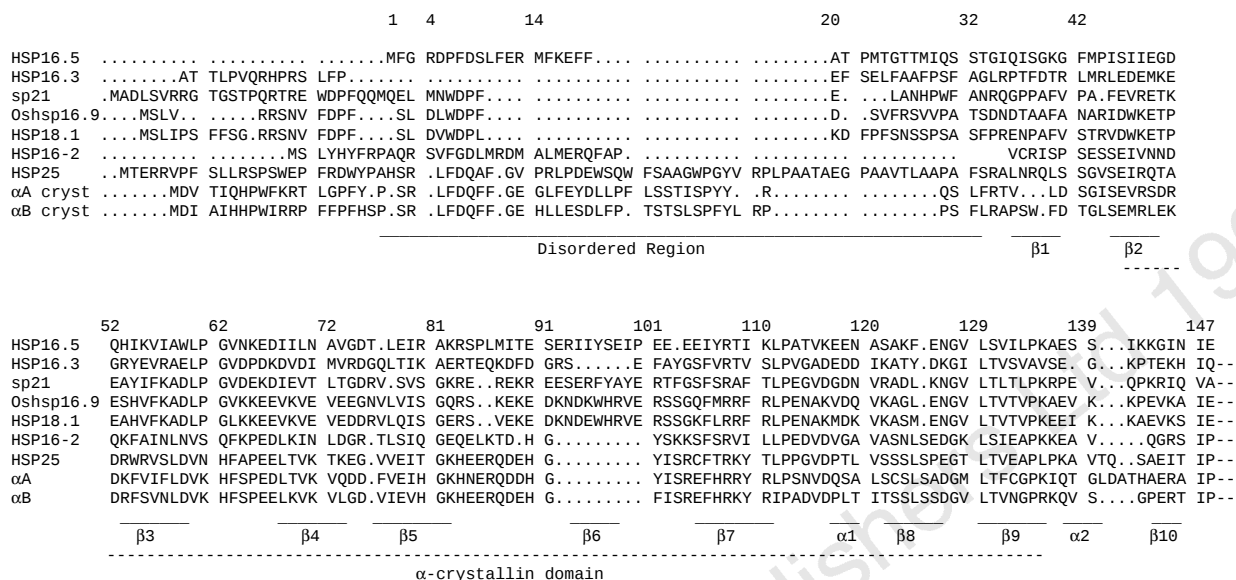
Physical Biosciences Division of the Lawrence Berkeley National Laboratory and the Department of Chemistry, University of California at Berkeley, 220 Melvin Calvin Laboratory, Berkeley, California 94720-5230, USA

The principal heat-shock proteins that have chaperone activity (that is, they protect newly made proteins from misfolding) belong to five conserved classes: HSP100, HSP90, HSP70, HSP60 and the small heat-shock proteins (sHSPs). The sHSPs can form large multimeric structures and have a wide range of cellular functions, including endowing cells with thermotolerance *in vivo*<sup>1,2</sup> and being able to act as molecular chaperones *in vitro*<sup>3–8</sup>; sHSPs do this by forming stable complexes with folding intermediates of their protein substrates<sup>9,10</sup>. However, there is little information available about these structures or the mechanism by which substrates are protected from thermal denaturation by sHSPs. Here we report the crystal structure of a small heat-shock protein from *Methanococcus jannaschii*, a hyperthermophilic archaeon. The monomeric folding unit is a composite  $\beta$ -sandwich in which one of the  $\beta$ -strands comes from a neighbouring molecule. Twenty-four monomers form a hollow spherical complex of octahedral symmetry, with eight trigonal and six square 'windows'. The sphere has an outer diameter of 120 Å and an inner diameter of 65 Å.

The sHSPs are abundant and ubiquitous in nature; they range in size from 12K to 42K and are found as large complexes of 200K–800K. The sHSPs share a sequence of about 100 residues which is homologous to  $\alpha$ -crystallin from the vertebrate eye lens, and is called the  $\alpha$ -crystallin domain or small-heat-shock-protein domain. The sHSP from *M. jannaschii* (MjHSP16.5, relative molecular mass 16.5K)<sup>11</sup> also contains an  $\alpha$ -crystallin domain composed of 90 residues (Fig. 1, residues 46 to 135). The domain has 20.7% sequence identity with human  $\alpha$ A-crystallin and 31.4% identity with rice HSP16.9 (refs 3, 12). The protein forms homogeneous oligomers and has molecular chaperone activity<sup>13</sup>. We have determined the crystal structure of MjHSP16.5 from *M. jannaschii* at 2.9 Å resolution by using single isomorphous replacement and non-crystallographic symmetry (NCS) averaging (Table 1 and Fig. 2).

MjHSP16.5 is a hollow spherical complex composed of 24 subunits generated by a three-fold crystallographic symmetry operation of an asymmetric unit containing eight subunits (Fig. 3a). These eight subunits, in turn, are related by three kinds of NCS: four two-fold, one three-fold, and one four-fold symmetries. Therefore, 24 subunits in the complex are related by an octahedral symmetry, with a total of twelve two-fold, three three-fold, and three four-fold NCS axes, and one three-fold crystallographic symmetry axis<sup>11</sup> (Fig. 3a). The outer diameter of the sphere is ~120 Å, and the inner diameter is ~65 Å (Fig. 3b). The inside of the sphere is hollow and no remarkable electron density is found at the current resolution. There are eight triangular and six square windows on the surface of the sphere.

Each folding unit is composed of nine  $\beta$ -strands in two sheets, two short  $3_{10}$ -helices, and one short  $\beta$ -strand (Fig. 4a). One of the  $\beta$ -strands comes from a neighbouring subunit. The amino-terminal 32 residues are highly disordered, but from residue 33 onwards, including the entire  $\alpha$ -crystallin domain (residues 46–135) and



**Figure 1** Sequence alignment of HSP16.5 with other small heat-shock proteins and  $\alpha$ -crystallins. The alignment was performed with 'pileup' in the GCG program package<sup>27</sup>. The sequences aligned are from *M. jannaschii* HSP16.5, *Mycobacterium tuberculosis* HSP16.3, *Stigmatella aurantiaca* sp21, rice

the carboxy-terminal extension, they are well ordered. The overall structure of the folding unit has dimensions of  $25 \text{ \AA} \times 28 \text{ \AA} \times 75 \text{ \AA}$ . The longest dimension is parallel to the length of the  $\beta$ -strands. Figure 1 shows the secondary structural elements of MjHSP16.5 aligned with homologous sHSPs. Two  $\beta$ -sheets are packed as parallel layers:  $\beta$ 1, 7, 5 and 4 of one subunit in one  $\beta$ -sheet and  $\beta$ 2, 3, 9 and 8 of the same subunit, and  $\beta$ 6 of a neighbouring subunit in the other  $\beta$ -sheet. In addition, there are two short  $3_{10}$ -helices ( $\alpha$ 1 and  $\alpha$ 2) and one short  $\beta$ -strand,  $\beta$ 10. The folding pattern of the monomer is similar to a domain of the immunoglobulin Fc fragment, although there is no sequence similarity between them.

Each subunit in the MjHSP16.5 complex makes extensive contacts with other subunits in the complex. There are three different subunit-subunit contacts: the contacts related by the two-, three-, and four-fold NCSs (Figs 4b-d and Table 2). The most extensive subunit intersubunit contacts are found around the two-fold axis: the backbone atoms of the  $\beta$ 2-strand in one subunit and the  $\beta$ 6-strand in the other subunit (in a two-fold related dimer) form hydrogen bonds to make an intersubunit composite  $\beta$ -sheet. There are also extensive hydrophobic contacts and ionic interactions as

Oshsp16.9, pea HSP18.1, *C. elegans* HSP16-2, murine HSP25, bovine  $\alpha$ A-crystallin, and bovine  $\alpha$ B-crystallin. The alignment is made only until the last amino acid of HSP16.5. The secondary structure of HSP16.5 was assigned using PROCHECK<sup>26</sup>, and the  $\beta$ -strands are labelled  $\beta$ 1- $\beta$ 10.

well as interbackbone hydrogen bonds that stabilize the two subunits related by two-fold symmetry (Fig 4b, e and Table 2). The putative substrate-binding region of the  $\alpha$ -crystallin domain of HSP18.1 which binds bis-ANS (ref. 10) is depicted in MjHSP16.5 (Fig. 4b) on the basis of the sequence alignment in Fig. 1 (loop between  $\beta$ 3 and  $\beta$ 4). In this loop, which is located outside the sphere, the conserved hydrophobic residues are involved in dimer interaction. Around the four-fold symmetry axis, four subunits make contact by hydrogen bonds, ionic, and hydrophobic interactions (Fig. 4c, e and Table 2). The C-terminal region (residues 142 to 147) of one subunit reaches out and interacts with  $\beta$ 4 and  $\beta$ 8 of a neighbouring subunit by hydrophobic interaction and backbone hydrogen bonding. One residue at the C terminus (Lys 141) is also involved in ionic interactions in this contact (Fig. 4e). Interactions among the subunits around the three-fold axis are the least extensive: one ionic interaction per subunit pair stabilizes two adjacent subunits at each corner of the triangular window (Fig. 4d, e and Table 2). Forty-one per cent of the solvent-accessible surface in each MjHSP16.5 monomer is buried at the intersubunit contacts ( $3,247 \text{ \AA}^2$  out of  $7,911 \text{ \AA}^2$ ). Of these, 48% of the contact surfaces contribute to dimer formation, 42% to tetramer formation, and only 10% to trimer formation. Of the dimer contacts, 66% are due to nonpolar interactions, whereas only 31% are nonpolar interactions in the tetramer. Combined with the composite nature of the folding unit described earlier, this suggests that the dimer may be the building block of the sphere.

The interior of the MjHSP16.5 spherical structure appears to be empty at the current resolution. The inside volume of the sphere is about  $140,000 \text{ \AA}^3$ , which is about 56% of the GroEL cylinder<sup>14</sup>. Thr 33, the first ordered residue in the subunit, is located inside the sphere and near the small square window around each four-fold symmetry axis. This suggests that the disordered N-terminal residues are inside the sphere, which is in agreement with the observation that the N terminus of HSP16-2 from *Caenorhabditis elegans* was found to be buried in the complex<sup>15</sup>. Interestingly, 49% of the solvent-accessible surfaces in the interior of the sphere are composed of nonpolar residues, in contrast to 22% on the outside surface, which is reminiscent of the GroEL structure<sup>14</sup>. This implies that the inside surface of the sphere is much more hydrophobic than the outside surface, although a part of this may be due to the areas

**Table 1** Crystallographic data

| Data collection and phasing statistics |                        |                       |          |
|----------------------------------------|------------------------|-----------------------|----------|
| Crystal                                | Native (CuK $\alpha$ ) | Native (NSLS)         | Se-Met   |
| Resolution ( $\text{\AA}$ )            | 30.0-3.2               | 30.0-2.9              | 30.0-3.2 |
| Completeness                           | 98.2%                  | 99.7%                 | 99.9%    |
| $R_{\text{sym}}$                       | 0.071                  | 0.037                 | 0.069    |
| $R_{\text{merge}}$                     |                        |                       | 0.11     |
| $R_{\text{cullis}}$ (%) <sup>*</sup>   |                        |                       | 0.78     |
| Phasing power†                         |                        |                       | 1.0      |
| Figure of merit (20.0-3.2)             |                        |                       | 0.258    |
| Refinement statistics                  |                        |                       |          |
| Resolution                             |                        | 15.0-2.9 $\text{\AA}$ |          |
| NCS-related monomers/asymmetric unit   |                        | 8                     |          |
| Unique reflections ( $F > 2\sigma$ )   |                        | 22,008                |          |
| Completeness ( $F > 2\sigma$ )         |                        | 80.6% (8.96)‡         |          |
| $R$ factor§                            |                        | 0.216 (0.251)‡        |          |
| Bond-length deviation from ideality    |                        | 0.012 $\text{\AA}$    |          |
| Bond-length deviation from ideality    |                        | 1.505 deg             |          |

<sup>\*</sup>  $R_{\text{cullis}}$  is the r.m.s. lack-of-error divided by the isomorphous difference.

† Phasing power is the mean  $F_o$  divided by the r.m.s. lack-of-error.

‡ Values for the test data aside to calculate the free  $R$  factor.

§  $R$  factor calculation was made from the data with  $2\sigma$  cutoff.

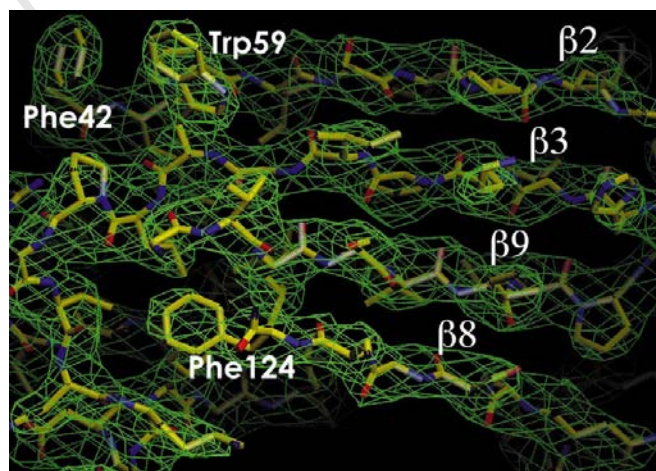
**Table 2 Intersubunit contacts in HSP16.5**

|                                                                                                                                                                                                                                                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Three-fold                                                                                                                                                                                                                                                        |
| A                                                                                                                                                                                                                                                                 |
| B Arg 93–Glu 78                                                                                                                                                                                                                                                   |
| C                                                                                                                                                                                                                                                                 |
| Four-fold                                                                                                                                                                                                                                                         |
| A 142O–72N, 142N–72O, 144N–70O, 145O–122N, 147N–122O                                                                                                                                                                                                              |
| B Asp 51–Arg 80, Lys 141–Glu 78                                                                                                                                                                                                                                   |
| C Ile 144–Leu 70, Ala 72, Leu 77, Ala 120, Ala 122, Val 131, Leu 133*;<br>Ile 146–Ile 68, Leu 70, Ala 122, Phe 124, Leu 129                                                                                                                                       |
| Two-fold                                                                                                                                                                                                                                                          |
| A. 45O–98N, 47N–96O, 47O–95N, 47O–96N, 49N–93O, 49O–93N, 126O–62N                                                                                                                                                                                                 |
| B. Glu 92–His 53, Glu 92–Lys 55, Glu 49–Arg 93, Asp 51–Arg 93                                                                                                                                                                                                     |
| C. Phe 42–Phe 42, Trp 59, Pro 44; Pro 61–Trp 59, Pro 61, Gly 127; Gly 62–Phe 42, Val 128;<br>Ile 86–Ile 48, Ile 57; Ile 88–Ile 48; Ile 94–Ile 48; Ile 95–Ile 35, Ile 47, Pro 112; Tyr 96–Ile 35,<br>Ile 37, Ile 45, Ile 47; Ile 99–Ile 57, Trp 59; Pro 100–Trp 59 |
| A, Backbone hydrogen bond.                                                                                                                                                                                                                                        |
| B, Side-chain ionic interaction or hydrogen bond.                                                                                                                                                                                                                 |
| C, Hydrophobic interaction.                                                                                                                                                                                                                                       |
| * Ile 144 in one monomer interacts with the listed residues in the neighbouring monomer.                                                                                                                                                                          |

covered by the disordered N-terminal residues.

There are eight triangular windows and six square windows with the edges of the window frames being 30 Å and 17 Å, respectively (calculated with the side-chain atoms beyond C $\beta$  deleted), on the surface of the hollow sphere of MjHSP16.5. Several negatively charged residues between  $\beta$ 5 and  $\beta$ 7 (Glu 90, Glu 101, Glu 103, and Glu 104) are located around the triangular windows. There are some charged residues (Asp 75, Lys 110, Glu 117, and Glu 118) around the square windows. These features are similar to those of the GroES structure, where there are two Glu residues around the orifice in the centre of the roof of the GroEL dome<sup>16</sup>. The size of the windows is large enough to allow small molecules such as enzyme substrates and products to diffuse in and out of the sphere. The triangular windows may be wide enough to allow even extended peptide chains to thread through them.

As the  $\alpha$ -crystallin domain of MjHSP16.5 has a well ordered folded structure and is involved in subunit contacts, oligomeric complexes are probably due to the  $\alpha$ -crystallin domain. The hydrophobic residues in the  $\alpha$ -crystallin domain that are involved in maintaining the tertiary structure of the folding unit are relatively well conserved among all species (Fig. 1). This is consistent with the finding that the predicted secondary structures and hydrophobic profiles of the  $\alpha$ -crystallin domains of the sHSP family (animal, plant and bacteria) are almost identical to those of MjHSP16.5 (ref. 17). However, some of the residues involved in subunit contacts are

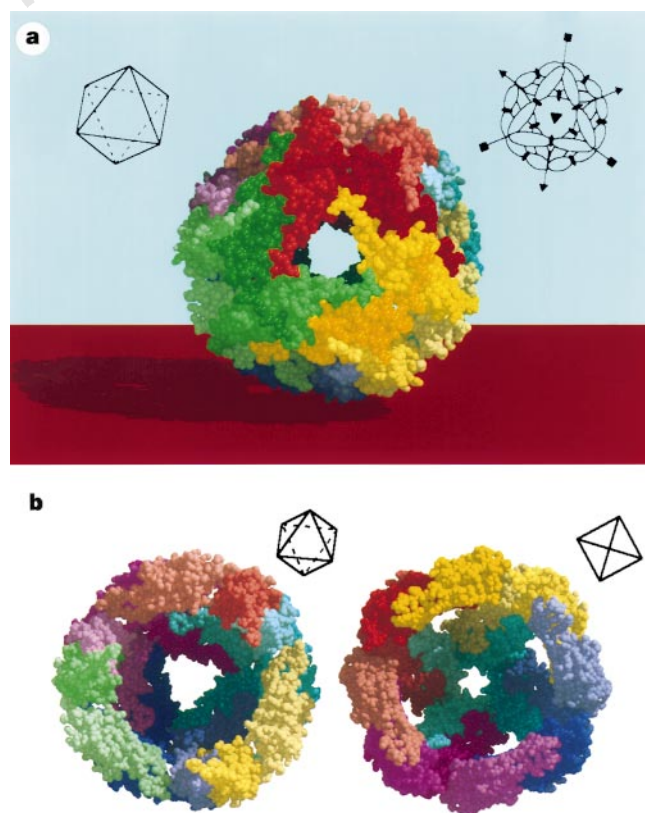


**Figure 2** Representative region of the experimental electron density map contoured at 1.2 $\sigma$ . One  $\beta$ -sheet region and three hydrophobic residues are labelled.

not conserved among all sHSPs (Fig. 1 and Table 2), suggesting that the size and the symmetry of oligomers may vary among different sHSPs although the unique fold of the sHSP monomer is conserved.

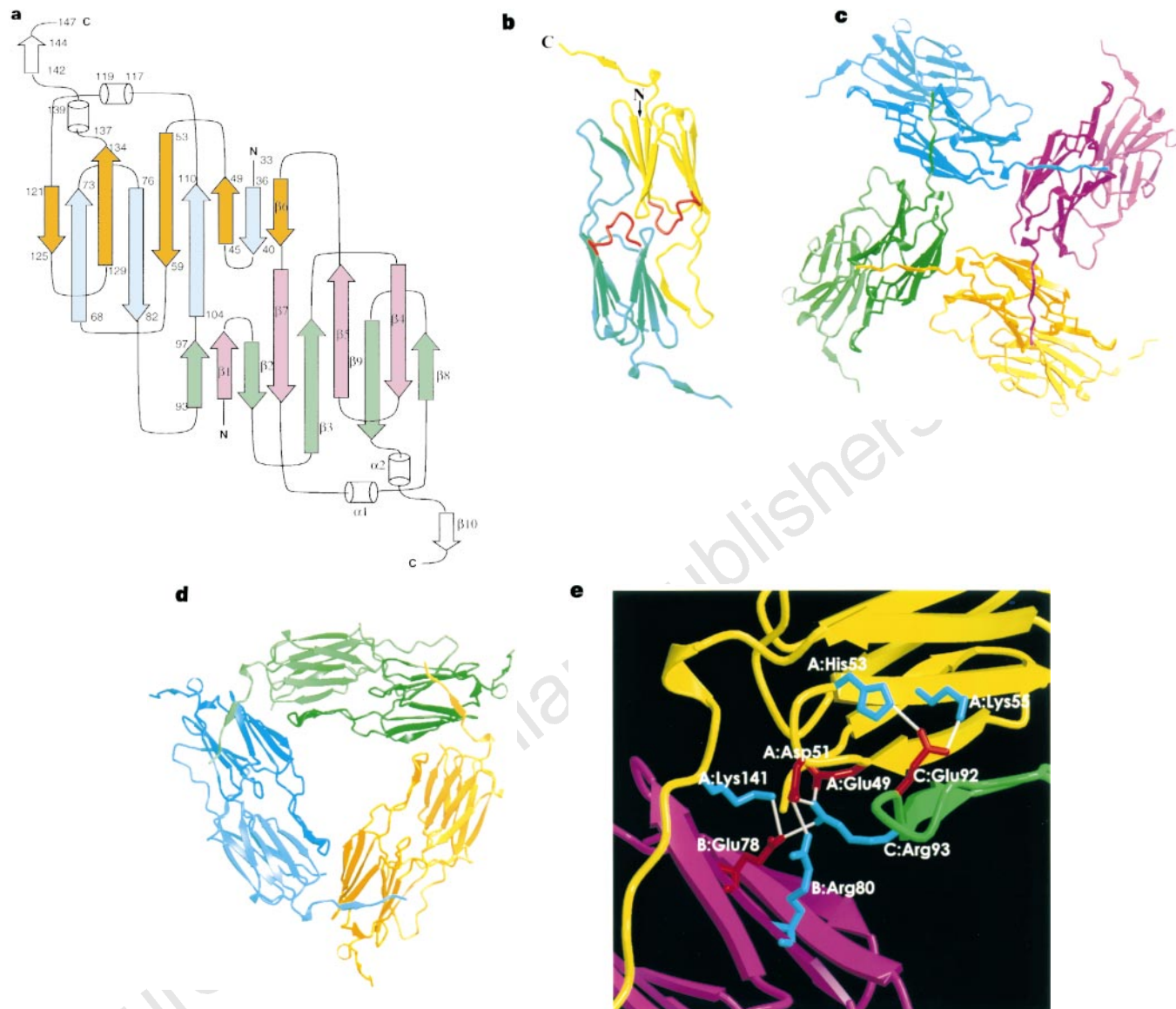
Among the diverse physiological functions of sHSPs, their *in vitro* chaperone activity appears to be common to most sHSPs. The assumption that the highly conserved  $\alpha$ -crystallin domain may be important for chaperone activity is contradicted by the observation that *Escherichia coli* expressing a rice protein deletion mutant, Oshsp16.9, fused to glutathione-S-transferase, where the C-terminal two-thirds of the  $\alpha$ -crystallin domain is missing, is protected from heat shock<sup>18</sup>. Another related observation is that mutations within the phenylalanine-rich region of  $\alpha$ B-crystallin, located N-terminal to the  $\alpha$ -crystallin domain, abolish chaperone activity *in vitro* without altering the size of their oligomeric complex<sup>19</sup>. The N-terminal regions seem to be necessary for oligomerization because the minimal  $\alpha$ -crystallin domain alone fails to form oligomers and has no chaperone activity *in vitro*<sup>20,21</sup>. Taken together, these observations seem to suggest that, although the  $\alpha$ -crystallin domain is important in oligomeric-complex formation, the N-terminal residues before the  $\alpha$ -crystallin domain are necessary not only for complex formation but also for chaperone activity.

There are many possible mechanisms to explain sHSPs' ability to protect other proteins from denaturation. One is that during the process of *in vivo* assembly of the hollow spheres, certain proteins or RNAs critical for the cells' survival under stress may get trapped in or on the outer surface of the spheres. Another is that the oligomeric



**Figure 3** Overall structure of sHSP. **a**, A space-filling model of the hollow sphere viewed along the crystallographic three-fold axis. Each HSP16.5 tetramer is represented in one colour with different shadings. Top right, schematic of the 24 subunits, drawn as ovals, and their symmetry elements; top left, octahedral symmetry of sHSP. The holes in the shadow of the sphere are generated by the triangular and square windows. **b**, The interior of the sphere is viewed along the three-fold axis (left) and the four-fold axis (right). The front one-third of each sphere is cut off to reveal the inside of the hollow sphere. Colour code as in **a**. An octahedron oriented in the same way as the corresponding sHSP is shown at the top.





**Figure 4** Subunit interaction of sHSP. **a**, Topology of the secondary structure of a MjHSP16.5 dimer. The first and last residue numbers for each secondary-structural element are indicated in the top (left) monomer; the secondary-structural elements are labelled in the monomer at the bottom (right). The first  $\beta$ -sheet of the top monomer is in blue and the second  $\beta$ -sheet is in yellow, but  $\beta 6$  (also yellow) is from the adjacent subunit. The first and second  $\beta$ -sheets of the bottom monomer are also shown in different colours (green and pink, respectively). **b**, Ribbon diagram of an HSP16.5 dimer viewed along the non-crystallographic two-fold symmetry axis. The N and C termini are indicated. The

form of sHSP is not necessary for chaperone or other stress-related activity, but is a storage state for sHSPs from which they can be disassembled quickly in response to the recurrence of external stress. □

## Methods

**Crystallization and data collection.** MjHSP16.5 was purified and crystallized as described<sup>11</sup>. Selenomethionine-substituted protein was crystallized under the same conditions. X-ray data of the native and selenomethionine-substituted crystals were collected at 3.2 Å resolution on a Rigaku R-Axis IIC imaging plate system. Another native data set was collected at 2.9 Å resolution at the beamline X12C at the National Synchrotron Light Source, Brookhaven (NSLS). The native and selenomethionine derivative data were processed and integrated by DENZO and scaled by SCALEPACK<sup>22</sup>. There are eight subunits in an asymmetric unit of the unit cell in the R3 space group with a  $V_m$  of 2.2 Å<sup>3</sup> dalton<sup>-1</sup>.

bis-ANS binding sites in the  $\alpha$ -crystallin domain of HSP18.1 (ref. 10) are shown in red. **c**, Four dimers related by a non-crystallographic four-fold symmetry. **d**, Three dimers related by a non-crystallographic three-fold symmetry. **e**, Several residues that are involved in ionic interaction in subunit contacts (Table 2) are shown. Subunits A (yellow) and B (magenta) are related by four-fold symmetry, B and C (green) by three-fold symmetry, and A and C by two-fold symmetry. Basic residues are coloured blue and acidic residues are coloured red. Possible ionic interactions are represented as white lines.

**Structure determination and refinement.** The first selenium site was found in the difference Patterson map. The other seven selenium sites were found in the difference Fourier map by using the phase calculated from the first selenium site at 3.2 Å. Only eight out of 40 selenomethionines in an asymmetric unit were found. The initial model was built on the map calculated from the eight selenomethionine sites and modified by solvent flattening with a solvent content of 45%. Six NCS operators were found from the initial model: four two-fold, one three-fold, and one four-fold rotations. The phases were improved further by averaging over eight NCS-related subunits in an asymmetric unit using the program DM<sup>23</sup>. At this stage, the electron density showed a single polypeptide chain per monomer without any disconnectivity. Amino acids were assigned from residues 33 to 147 using program O<sup>24</sup>. The first 32 amino acids were unassignable because of disorder. The other seven subunits were generated from the first subunit model by the NCS operators.

Several cycles of rigid body refinement, positional refinement, and simulated



annealing with tight non-crystallographic restraints in X-plor<sup>25</sup> were performed at 3.2 Å resolution. The refinements were continued by using the native data at 2.9 Å resolution collected at the NSLS and resulted in an *R* value of 21.6% and a free *R* value of 25.1% with the bulk solvent correction and B-factor refinement. The refined model included eight subunits, each containing 116 residues from amino acids 33 to 147. The root-mean-squares deviation for all atoms among the eight subunits in an asymmetric unit is 0.041 Å. Most (85.2%) of the non-glycine amino acids were in the most favourable regions, and the remaining 14.8% were in the allowed region in the Ramachandran plot drawn by PROCHECK<sup>26</sup>.

Received 29 April; accepted 9 June 1998.

- Schirmer, E. C., Lindquist, S. & Vierling, E. An *Arabidopsis* heat shock protein complements a thermotolerance defect in yeast. *Plant Cell* **6**, 1899–1909 (1994).
- van den IJssel, P. R., Overkamp, P., Knauf, U., Gaestel, M. & de Jong, W. W. Alpha A-crystallin confers cellular thermoresistance. *FEBS Lett.* **355**, 54–56 (1994).
- Horwitz, J. Alpha-crystallin can function as a molecular chaperone. *Proc. Natl Acad. Sci. USA* **89**, 10449–10453 (1992).
- Jacob, U., Gaestel, M., Engel, K. & Buchner, J. Small heat shock proteins are molecular chaperones. *J. Biol. Chem.* **268**, 1517–1520 (1993).
- Merck, K. B. *et al.* Structural and functional similarities of bovine alpha-crystallin and mouse small heat-shock protein. A family of chaperones. *J. Biol. Chem.* **268**, 1046–1052 (1993).
- Lee, G. J., Pokala, N. & Vierling, E. Structure and *in vitro* molecular chaperone activity of cytosolic small heat shock proteins from pea. *J. Biol. Chem.* **270**, 10432–10438 (1995).
- Chang, Z. *et al.* Mycobacterium tuberculosis 16-kDa antigen (Hsp16.3) functions as an oligomeric structure *in vitro* to suppress thermal aggregation. *J. Biol. Chem.* **271**, 7218–7223 (1996).
- Jaenicke, R. & Creighton, T. E. Junior chaperones. *Curr. Biol.* **3**, 234–235 (1993).
- Ehrnsperger, M., Graber, S., Gaestel, M. & Buchner, J. Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. *EMBO J.* **16**, 221–229 (1997).
- Lee, G. J., Roseman, A. M., Saibil, H. R. & Vierling, E. A small heat shock protein stably binds heat-denatured model substrates and can maintain a substrate in a folding-competent state. *EMBO J.* **16**, 659–671 (1997).
- Kim, K. K., Yokota, H., Kim, S.-H. Purification, crystallization and preliminary X-ray crystallographic analysis of small heat shock protein homolog from *Methanococcus jannaschii*. *J. Struct. Biol.* **121**, 76–80 (1998).
- Tseng, T. S. *et al.* Two rice (*Oryza sativa*) full-length cDNA clones encoding low-molecular-weight heat-shock proteins. *Plant Mol. Biol.* **18**, 963–965 (1992).

- Kim, R., Kim, K. K., Yokota, H. & Kim, S.-H. Small heat shock protein of *Methanococcus jannaschii*, a hyperthermophile. *Proc. Natl Acad. Sci. USA* (in press).
- Braig, K. *et al.* The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578–586 (1994).
- Leroux, M., Melki, R., Gordon, B., Batelier, G. & Candido, E. P. M. Structure–function studies on small heat shock protein oligomeric assembly and interaction with unfolded polypeptides. *J. Biol. Chem.* **272**, 24646–24656 (1997).
- Hunt, J. F., Weaver, A. J., Landry, S. J., Gierach, L. & Deisenhofer, J. The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature* **379**, 37–45 (1996).
- Caspers, G.-J., Leunissen, J. A. M. & de Jong, W. W. The expanding small heat-shock protein family, and structural predictions of the conserved “α-crystallin domain”. *Mol. Evol.* **40**, 238–248 (1995).
- Yeh, C. H. *et al.* Expression of a gene encoding a 16.9-kDa heat-shock protein, Oshsp16.9, in *Escherichia coli* enhances thermotolerance. *Proc. Natl Acad. Sci. USA* **94**, 10967–10972 (1997).
- Plater, M. L., Goode, D. & Crabbe, M. J. C. Effects of site-directed mutations on the chaperone-like activity of αB-crystallin. *J. Biol. Chem.* **271**, 28558–28566 (1996).
- Merck, K. B. *et al.* Comparison of the homologous carboxy-terminal domain and tail of alpha-crystallin and small heat shock protein. *Mol. Biol. Rep.* **18**, 209–215 (1993).
- Leroux, M., Ma, B. J., Batelier, G., Melki, R. & Candido, E. P. M. Unique structural features of a novel class of small heat shock proteins. *J. Biol. Chem.* **272**, 12847–12853 (1997).
- Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, UK, 1993).
- Cowtain, K. D. & Main, P. Improvement of macromolecular electron-density maps by the simultaneous application of real and reciprocal space constraints. *Acta Crystallogr. D* **49**, 148–157, 1993.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *X-PLOR version 3.1* (Yale Univ. Press, New Haven, CT, 1993).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Genetics Computer Group *Program manual for the GCG package, version 7* (GCG, Madison, WI, 1991).

**Acknowledgements.** We thank D. King for doing the electrospray mass spectrometry; H. Yokota for help with protein preparation, D. Boisvert for critical discussion of the manuscript, and R. Sweet at the NSLS for data collection. This work was funded by the US Department of Energy (R.K. and S.-H.K.).

Correspondence and requests for materials should be addressed to S.-H.K. (e-mail: SHKIM@LBL.GOV). The identity code of the coordinates and structure factors are 1SHS and 1SHSSF, respectively, in the Brookhaven Protein Data Bank.

# YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

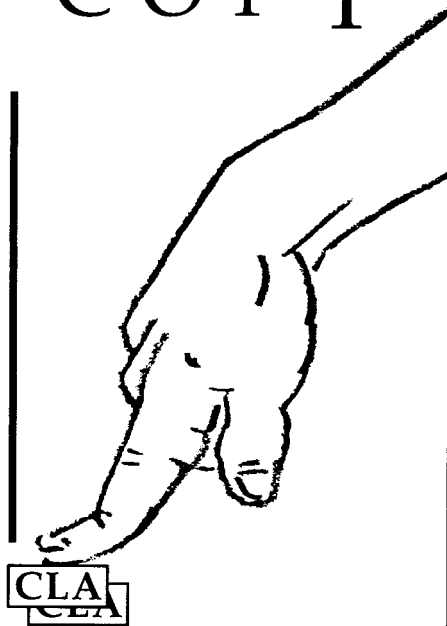
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised infringing photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited  
90 Tottenham Court Road, London W1P 0LP  
Telephone: 0171 436 5931  
Fax: 0171 436 3986

# Scientists with business flair in demand

**Despite the economic crisis in the Pacific rim of Asia, job opportunities for scientists, in particular ethnic Chinese scientists, are growing. Demand for those with business experience is strong as countries invest in applied research and development to try to climb their way out of their difficulties.**

**This Careers and Recruitment feature was written by David Swinbanks and Richard Nathan of Nature's Tokyo office.**

In the early 1990s, waves of Chinese scientists returned to Taiwan and Hong Kong from the West to positions in universities and government research institutes (see *Nature* 383, 11; 1996). The 'golden era' for academics in these two economies is now largely over, and there are far more applicants than posts. But opportunities for scientists are opening up in business and in some areas of research, for example, in the application of biotechnology to agriculture, fisheries and medicine.

Taiwan is continuing to build science parks that are a major source of employment. And government leaders and academics in China and Hong Kong are discussing setting up science parks in the region between Guangzhou and Hong Kong.

The next few years may be the best time for Chinese scientists to return to China. The Chinese Academy of Sciences has just announced a major recruitment drive to bring back 600 Chinese scientists as the academy reorganizes for the next century (see box above right). Prospects look particularly bright for those working in genomics. Centres for human genome research are being set up in Shanghai and Beijing, with substantial financial support from the government as well as from state-owned companies, and there are also several initiatives by Western companies in this area (see page 602).

## Commercial opportunities

China has one-fifth of the world's population and vast natural genetic resources, so it is not just government officials and scientists

## New blood for Chinese academy

"Now is a golden opportunity for Chinese scientists to return to China," says Chunli Bai, a vice-president of the Chinese Academy of Sciences, explaining the academy's plans dramatically to reorganize its 123 institutes over the next 13 years. The number of institutes and the 68,000 permanent employees of the academy will be drastically reduced by the year 2010, but 600 scientists will be recruited over the next three years (see *Nature* 394, 7; 1998).

"We are looking for bright young people under 45, most of them from overseas. We want people who are good in science but also people who have good management skills, for example, people from industry experienced in corporate management," Bai explains. "We expect about half to be pure scientists, but that is not a precise figure."

The recruits will be offered accommodation, equipment, graduate students and 2 million yuan (US\$240,000) in

start-up research funds. They will be given the title 'research professor', which will entitle them to supervise graduate students. The positions will be long term on a 'tenure track' system. Scientists will be assessed every four years and if they fail to meet standards they will be "out", says Bai. Only after passing three such assessments (after 12 years) will lifetime employment be offered.

To compensate for the academy's rather low salaries, the recruits will be given performance-based bonuses. "We will take care of them," says Bai, and give them "more and more support".

There will also be many openings in the academy for 'mobile' researchers, such as graduate students, postdoctoral fellows and visiting professors. The academy hopes to more than double the number of such 'mobile' people to 10,000 in the next three years, and to further boost the number to 30,000 by 2010.

who are interested in genomics. "We see great commercial opportunities," says Paul Watkins of the US pharmaceutical company Axys, which has set up a joint venture in Shanghai — Shanghai GeneCore BioTechnologies — with Perkin Elmer Applied Biosystems and SiniWest Holdings. This consortium is searching for genes involved in liver cancer under a contract with the Chinese Ministry of Science and Technology, but it expects to branch out into other areas, including applications of biotechnology to agriculture and fisheries.

The US agricultural giant Monsanto has just obtained permission from the Chinese

authorities for the commercial release of its genetically modified bollworm-resistant cotton. Bollworm has devastated cotton output in China, the world's leading producer of this crop. Monsanto is collaborating with the new Institute of Molecular Agrobiology in Singapore (see page 604).

In Taiwan too, government science officials are abuzz with talk of the need to develop agrobiotechnology (see page 603).

Scientists with experience in business are particularly sought after. "We want people trained in the West with corporate management skills as well as a sound scientific background," says Ming-Wei Wang, chairman

## Tsinghua University recruits for excellence

Tsinghua University in Beijing is one of China's leading universities. It has a much stronger reputation in engineering than in science — but that is about to change. In the 1950s, Chinese universities were forced to specialize, following the model of the former Soviet Union. As a result, Tsinghua lost its five science departments to neighbouring Peking University. Now, however, a new science building will open shortly on Tsinghua's huge campus of 50,000 faculty members, their families, students and support staff.

In the new building is a Centre for Advanced Study modelled on the Princeton

Institute for Advanced Study in the United States. The centre will focus on theoretical studies of basic sciences, and be led by Hwa Tung Nieh, who has recently returned to China from the State University of New York at Stony Brook. There will be a core of seven or eight professors who will be "very, very carefully selected", says Nieh. Twelve associate professors will be hired on five-year contracts. "In principle" they will have to leave after five years, although they will probably get positions elsewhere in the university. Bin-lin Gu, a vice-director of the centre and dean of science, hopes the centre

will become an "incubator" of high-quality scientists in the university.

There will be 20 postdoctoral positions in mathematics, theoretical physics, chemistry and possibly biology. Nieh expects to get many recruits from overseas, in particular Europe, as well as from local universities. The centre's faculty members will receive supplements to double their salaries compared with other university staff, and housing will be provided. "In five years, we should have a reasonable centre," says Nieh. "In 10 or 20 years, as China's economy grows, we hope to become a great centre."

## Openings in human genome research in China

Three new centres for human genome research in Shanghai and Beijing offer about 100 positions for genome researchers and more genome initiatives are expected soon from the Chinese government (see *Nature* 394, 109; 1998).

**Shanghai Human Genome Centre.** Headed by Zhu Chen of Shanghai Second Medical University. Will have 35–40 researchers in a year's time, divided into four groups: expression profiles; genotyping of disease; single nucleotide polymorphisms; and functional genomics. Liver cancer, which

is prevalent in Shanghai, is a particular focus. Start-up funds of 60 million yuan (US\$7 million) are coming from local and central government.

**Beijing Human Genome Centre.** Headed by Boqin Qiang, vice-president of the Chinese Academy of Medical Sciences. Will have about 30 researchers in three divisions, devoted to sequencing; genetic resources; and bioinformatics. There will be a focus on cardiovascular disease and diseases of the nervous system. Start-up funds of nearly 100 million yuan (\$12 million) are coming from

central and local government and a state-owned real estate company.

**Shanghai GeneCore BioTechnologies.** A private sector initiative by three US companies — PE Applied Biosystems, Axys Pharmaceuticals, and Siniwest Holdings — which will have about 30 staff within a year. This joint venture in collaboration with local institutions is hunting for genes that cause liver cancer but expects to branch out into other areas of genomics. The three companies will make an initial investment of \$10 million.

and general manager of Shanghai GeneCore BioTechnologies.

As there is state support for research with a socioeconomic impact, such scientists-turned-businessmen will be in increasing demand. Recognizing this, members of the Overseas Chinese Physics Association (OCPA) last November formed the Chinese Association for Science and Business (CASB) with US-based Chinese professionals trained in science and engineering but working in finance, computer and information technology, and other business sectors. The general secretary of CASB is Ning Luo, of the department of physiology and biophysics at Mount Sinai School of Medicine in New York.

Luo says: "We see growing opportunities in mainland China as well as Hong Kong, Taiwan and Singapore, not only for Chinese scientists, but also for an even larger number of researchers-turned-entrepreneurs, business executives and professionals. On the one hand are the economies and markets which are ripe for extensive infusion of technology products and investment, and on the other hand is a large pool of expatriate scholars in North America, Europe and Japan and other advanced countries, supplied every year by the world's largest overseas student group.

"Because of their earlier economic successes, Hong Kong, Taiwan and Singapore have had much more experience than mainland China in attracting and utilizing returned talents," explains Luo. In this regard, CASB and OCPA, with the National Natural Science Foundation of China, are organizing an international conference in Guangzhou near Hong Kong this month to explore the "untapped mainland China market for scientists and engineers trained overseas", says Luo. For details of CASB and the conference, see [www.casbi.aan.net](http://www.casbi.aan.net).

Another network of Chinese expatriates is the North American Chinese Association of Science and Technology, established in Boston in 1994 ([www.voicenet.com/~yjinnacast](http://www.voicenet.com/~yjinnacast)). This has more than 3,000 members including scientists and professionals in

finance, health care, engineering and education. A chapter of the association formed a corporation with the Pudong New Area in Shanghai last January. This body, the Pudong-Princeton Consulting Corporation, aims to help develop health care and telecommunications industries in the Pudong area, which is also the base for one of China's new human genome research centres (see [www.voicenet.com/~yjinnacast](http://www.voicenet.com/~yjinnacast)).

### Poor salaries

The very low level of salaries for scientists in the public sector in China remains a major disincentive for people to return. Even university professors in richer departments can only expect about 2,500 yuan (US\$300) a month. In some fields, such as physics, salaries are often much lower.

Mechanisms are being devised to supplement incomes with performance-based bonuses and to provide accommodation and large start-up grants to scientists returning to the country.

Zihe Rhao, for example, who returned to China from Oxford University in the United Kingdom, was given about US\$0.5 million in support from Tsinghua University to set up a structural biology laboratory. And university faculty members can now supplement their incomes from government project grants, provided they work in approved areas of socioeconomic impact.

The new genome research centres in Beijing and Shanghai are being set up jointly by universities, the academy of sciences, and local and central government, and will have new employment practices. Most researchers will be employed on contracts rather than on a permanent basis, and will be paid performance-based salaries. Similarly, a new centre for advanced studies at Tsinghua University will employ most of its researchers on five-year contracts (see page 601).

Although scientists can get higher salaries in industry, the business environment on the Chinese mainland is still "far from ideal", says Luo. But the situation is changing. "The Chinese leadership recently have empha-

sized more and more the importance of science, technology and education in the next phase of economic development," he says. "Especially since the onset of the Asian economic crisis, people realize that labour-intensive export-oriented companies, which have been driving the Chinese economy in the last decade, are not sustainable." The economic crisis has provided the "impetus" to develop technology industries, he says. This in the long run will lead to a "much more attractive environment" for overseas scientists.

### Short-stay visits

For those Chinese scientists still disinclined to return, there are growing opportunities for short-term visits. The National Natural Science Foundation of China has since 1992 provided more than 200 small project grants a year, worth about 13,500 yuan (\$1,600) each, for research in China lasting a month or two. This year, the foundation has introduced a scheme that provides much larger grants of about 300,000 yuan (\$36,000) for two to three months for 30 or 40 overseas scientists.

Along similar lines, the municipal government of Shanghai has created the Magnolia Foundation to provide monthly stipends for visiting scientists. Taiwan's National Science Council now offers hundreds of postdoctoral fellowships and opportunities for visiting scholars with generous salaries and allowances to attract people back for short- or long-term visits (see page 603). And Singapore has introduced a scheme to attract top scientists to work on a part-year basis for several years (see page 604).

With the exception of Singapore and Hong Kong, there are at present comparatively few job opportunities in the region for non-Chinese scientists. Academia Sinica in Taiwan is beginning to open up positions to other nationalities, and Singapore is very open in its recruitment policies, drawing scientists from Europe, North America, Australasia, India and China. But in the region as a whole the overwhelming number of jobs are for those of ethnic Chinese origin.

# Taiwan aims to become sci-tech island

In Taiwan, an overseas education is highly valued. Unlike many Western countries, where senior government officials and politicians are often trained lawyers, many senior bureaucrats and almost all the government ministers in Taiwan have PhDs from overseas universities, mostly in science and engineering. Science and technology is an increasingly important government priority.

The country's research and development budget is growing by 10 per cent a year, despite the fact that overall government spending is only increasing by 3 per cent. The government is determined to build Taiwan into a 'sci-tech island' by increasing funding and improving the quality of its researchers.

Last December, Taiwan published its first white paper on science and technology. Later this year legislators at the Executive Yuan will debate a science and technology basic law, which will make it easier to recruit overseas nationals and simplify regulations for intellectual property rights. Political support for science and technology reaches to the top of government. Taiwan's prime minister recently set up a science and technology committee, which he chairs, to improve policy implementation. The deputy prime minister, Chao-Shiuan Liu, is the former chairman of the National Science Council (NSC).

## Political backing for recruitment

Opportunities are opening up for researchers, particularly ethnic Chinese, as a result of this political momentum, increased funding, and a series of new national programmes, including a three-year agricultural biotechnology programme launched in July. There are also now more opportunities for non-Chinese postdoctoral fellows, especially at Academia Sinica, Taiwan's national academy of sciences.

Taiwan has become attractive academically to newly graduated researchers and to more senior people, according to Chao-Han Liu, president of National Central University, who returned from the University of Illinois in 1990. The NSC supports newly returned researchers by making it relatively straightforward for them to win research funding. After only a few years it is possible to win grants of up to US\$0.5 million to equip laboratories. "It is a once in a lifetime opportunity," says Liu. A 'frontiers sciences' programme launched last year selects five or six researchers a year from the 3,000 life science grant holders. Each receives a five-year grant of NT\$3.5–8 million a year (US\$100,000–200,000). Among the first five holders is a 36-year-old at Academia Sinica who recently returned from the United States.

According to young researchers in Taiwan, facilities are comparable if not better

than those at universities in the United States. There is also less pressure to bring grant money into the department, so people can concentrate more on their research.

From 1987 to 1993 was a golden period for Taiwan's universities: funding increased and universities built new facilities and departments. Many positions became available in engineering and information technology. But overall university funding is now flat and opportunities in these fields, where Taiwan is now relatively strong, are more competitive. One vacancy receives up to 130 applications. But funding for university research is now a priority and is increasing by 15 per cent a year. "We need to beef up university research and recruit the best minds," says Steve Hsieh, vice-chairman of the NSC.

A shift in opportunities towards the life sciences is also taking place as new institutes are set up and universities expand their life science departments. Opportunities are also arising in the social sciences and research into sustainable development. National Central University, for example, a medium-sized comprehensive university which is trying to establish itself as one of Taiwan's key research orientated universities, set up a life science graduate school five years ago. This year, it received permission to start an undergraduate programme and will recruit 15–20 new faculty members over the next five years.

Academia Sinica has been a magnet for senior researchers returning to the region. Many of the heads of its 23 institutes are scientists returned from overseas. This trend is set to continue, with senior staff retiring, new institutes planned and general expansion at the academy. This year, the preparatory office of the Institute of Biological Agricultural Sciences was set up and a new Centre for Applied Science and Engineering is awaiting final approval. Eight people have been recruited for the Institute of Biological Agricultural Sciences and it will recruit a fur-

ther eight or so principal investigators every year for the next few years, according to Shang Fa Yang, academy vice-president.

The national agricultural biotechnology programme launched last month will fund at least 60 new projects, according to Yang. This, Yang believes, will force universities to recruit more faculty members and will lead to a 20 or 30 per cent annual increase in the number of people doing biotechnology related research in Taiwan.

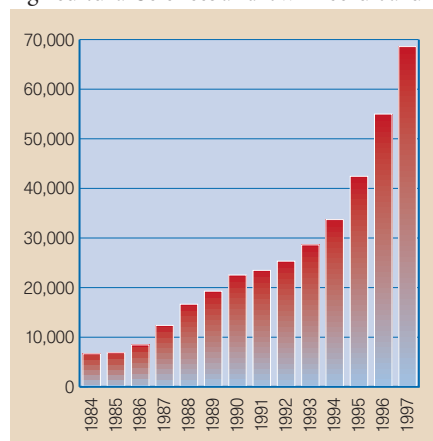
## Investing in health

The National Health Research Institutes (NHRI) opened in January 1996. According to its director, Cheng-Wen Wu, it will be the equivalent of the US National Institutes of Health. The new organization has been ferociously recruiting staff from overseas. More than half of its planned 10 divisions have been set up and it plans to double in size to 400 or 500 over the next five years.

In January this year, Ming-Chu Hsu joined the NHRI as director of the division of biotechnology and pharmaceutical research. She was previously the research director of the department of oncology at Hoffmann-La Roche in the United States. She is building a research team and is excited by the prospects ahead. She plans to recruit 20–25 PhDs over the next two years to work in an "integrated drug-discovery group". Hsu is also looking for experienced project managers. Her team will concentrate on three areas: chemical pharmaceuticals; diagnostics including biochips and biosensors; and biological pharmaceuticals.

The NSC-administered science-based industrial park in Hsinchu has created 70,000 jobs (see graph), and has helped to make Taiwan a leader in the production of personal computers and integrated circuits. Following this success, a new science park will open later this year in Tainan, in the south of Taiwan. A third science park is already being planned, and extensions to the Hsinchu park are also proposed, to concentrate on biotechnology related industries.

The NSC has a series of programmes to attract overseas researchers to Taiwan (see [www.nsc.gov.tw](http://www.nsc.gov.tw)), including a postdoctoral programme with 500 positions available a year for terms of between three months and three years. These posts attract monthly salaries of NT\$50–60,000 (US\$1,450–1,750), year-end bonuses and air fares. The council also has a short-term visiting programme for 300 scholars a year and an 'honourable visitor programme' through which it invites 40 senior scientists to promote international collaboration. Distinguished visiting scientists receive return air tickets and an honorarium of up to US\$600 a day.



On the up and up: growth in jobs at the science-based industrial park in Hsinchu.



# Singapore attracts foreign talent

Attracting "foreign talent" to Singapore is a high-profile government policy. In addition to its population of almost 3 million, there are more than 300,000 foreign workers in Singapore, including many highly educated people. "Science and technology people are particularly welcome," according to Chong Lit Cheong, executive director of the National Science and Technology Board (NSTB).

Some locals are concerned about the competition for jobs. But government officials constantly stress publicly the importance of developing the right skills for the whole country's future benefit.

Despite economic uncertainty, Singapore is expanding its universities and setting up new research institutes. It plans to double the number of postgraduate students at universities by the end of the century, and wants to develop into the education hub of the region. It aims to build a pool of qualified people to power its industry into the next century and to attract multinational companies to set up research and development centres on the island.

## Critical mass

The Institute of Molecular and Cell Biology (IMCB), set up in 1987, has helped to put Singapore on the international research map. It has attracted many high-calibre researchers, expanding from a group of 38 scientists to a research staff of 250. Its main research areas include cell regulation, functional genomics, immunology, virology and drug discovery, and its researchers have built an impressive publication record by any standards. "Singapore is trying to develop a credible critical mass of people," says Chris Tan, director of the institute.

In the wake of IMCB's success, a string of institutes have been founded in Singapore, including the Institute of Molecular Agrobiolology (IMA) established in 1995; the Bioinformatics Center in 1996; the Institute of Materials Research and Engineering in 1997; the Cancer Therapeutics Research Group this year; and the Singapore Synchrotron Light Source, which should be operational by next year. The IMA has just moved into a new building on the campus of the National University of Singapore and has 10 groups working on basic and applied research. It hopes to expand its research staff from 150 to 250 over the next three years, and is looking for outstanding candidates. Many other institutes as well as the IMCB are currently looking for recruits.

According to government figures, there are about 4,000 research scientists in Singapore (excluding technicians), of whom 40 per cent are foreign. One of the NSTB's main



Looking to recruit 100 'outstanding' staff: Singapore's Institute of Molecular Agrobiolology.

objectives is to encourage the development of Singapore's human resources and the recruitment of overseas researchers for industry and the research institutes. "If you have the right skills you are always welcome," says Cheong. The board has four schemes targeted at recruiting researchers from overseas. Researchers looking for work in Singapore should send their CVs, indicating their areas of interest, to the NSTB database ([www.nstb.gov.sg/career](http://www.nstb.gov.sg/career)). The board matches candidates with companies and institutes.

The board also has a foreign researchers' programme to assist companies in the recruitment of experienced foreign researchers, providing 50 per cent funding of recruitment by helping with relocation costs, salaries and housing allowances for up to two years. There are also postdoctoral and post-masters' fellowships to attract foreign researchers to the universities and institutes in Singapore. These researchers, mainly from mainland China, are expected to move on to industry after two years.

## Magnet for talent

The economic crisis in the region has meant job cuts for production staff at semiconductor manufacturers in Singapore, but there is still a shortage of highly qualified people, especially in information technology. Many pharmaceutical companies, including GlaxoWellcome, Becton-Dickinson, Baxter and Rhone-Poulenc Rorer, have research programmes in Singapore, and biotechnology and life-science companies are also setting up facilities. Perkin-Elmer set up a plant for DNA thermal cycles, machines that amplify DNA using PCR technology, in Singapore this year.

The NSTB launched a 'magnet' pro-

gramme called the Temasek Professorship last year. Temasek is the ancient name for Singapore in Malay. The idea is to target individuals in areas in which Singapore wants to develop, and hire them to head or start laboratories. These areas include biological sciences, mathematics, semiconductors and data storage. Singapore plans to recruit between 20 and 30 Temasek professors and has allocated a budget of S\$150 million (US\$89 million).

The professorships will be for between three and five years but the chosen individuals will be required to spend only 50 per cent of their time in Singapore. The first professor has already been recruited from the United Kingdom but the board says recruitment "is not an easy process". It has set up a steering committee to identify key people. The benefits on offer are substantial and, according to university officials, a whole team of researchers will be funded if necessary.

The Nanyang Technological University, which has recruited the first Temasek professor (a specialist in precision engineering), recruits 150 postdoctoral fellows every year ([www.ntu.ac.sg](http://www.ntu.ac.sg)). It has set up several new centres, according to Hong Siang Tan, director of research at the university. These include a biomedical engineering centre, launched in April last year, which has dozens of big projects in progress.

The National University of Singapore is planning a Tropical Marine Science Institute, according to Chang-Chieh Hang, the deputy vice-chancellor. The institute will be set up on St John's Island and will conduct research into marine biotechnology, oceanography, offshore engineering and tropical fish farming. The university is looking for a director for the institute, which will have between 150 and 200 research staff. □